From:	Roark, Jessica
Sent:	Wednesday, December 17, 2003 2:44 PM
To:	Delaval, Jan
Subject:	10/045180

Jan,

When things are up and running......

Please search, including pending, the following from 10/045180:

SEQ ID NO:3 against the DNA database
SEQ ID NO:3 as an oligo against protein

SEQ ID NO:6 against DNA
SEQ ID NO:6 as an oligo against protein

SEQ ID NO:4 against DNA
SEQ ID NO:5 against DNA

Results on paper please.

Thanks!

Jessica H. Roark
CM1 8A03
Mailbox 9E12
Art Unit 1644
703 605-1209

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 14:50:33 ; Search time 2689.5 Seconds
                                             (without alignments)
                                             1429.821 Million cell updates/sec

Title: US-10-045-180A-3
Perfect score: 514
Sequence: 1 MRTITLLSAFLLVALQAWAB........GEHLGGTCFILGERYPICCY 94

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext 0.5
               Ygapop 10.0 , Ygapext 0.5
               Fgapop 6.0 , Fgapext 7.0
               Delop   6.0 , Delext   7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10045180/runat_17122003_145023_12528/app_query.fasta_1.860
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10045180 @CGN_1_1_5283 @runat_17122003_145023_12528 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=30 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
          1: gb_ba:*
          2: gb_htg:*
          3: gb_in:*
          4: gb_om:*
          5: gb_ov:*
          6: gb_pat:*
          7: gb_ph:*
          8: gb_pl:*
          9: gb_pr:*
         10: gb_ro:*
         11: gb_sts:*
         12: gb_sy:*
         13: gb_un:*
         14: gb_vi:*
         15: em_ba:*
         16: em_fun:*
         17: em_hum:*
         18: em_in:*
         19: em_mu:*
         20: em_om:*
         21: em_or:*
         22: em_ov:*
         23: em_pat:*
         24: em_ph:*
         25: em_pl:*
         26: em_pr:*
         27: em_sts:*
         28: em_un:*
         29: em_vi:*
         30: em_htg_hum:*
         31: em_htg_inv:*
         32: em_htg_other:*
         33: em_htg_mus:*
         34: em_htg_pln:*
         35: em_htg_rod:*
         36: em_htg_mam:*
         37: em_htg_vrt:*
         38: em_sy:*
         39: em_htgo_hum:*
         40: em_htgo_mus:*
         41: em_htgo_other:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	514	100.0	453	6	A98571	A98571 Sequence 2
2	514	100.0	453	6	BD074746	BD074746 Human def
3	393.5	76.6	4415	6	A98570	A98570 Sequence 1
4	393.5	76.6	4415	6	BD074745	BD074745 Human def
5	374	72.8	128544	9	AF238378	AF238378 Homo sapi
6	300	58.4	163093	2	AC116559	AC116559 Papio anu
7	300	58.4	188604	2	AC116558	AC116558 Papio anu
8	227	44.2	542	9	HSCCSHP4	X65977 H.sapiens m
9	225.5	43.9	347	9	AF188268	AF188268 Macaca mu
10	225.5	43.9	435	9	AF184160	AF184160 Macaca mu
11	223.5	43.5	433	9	AF184159	AF184159 Macaca mu
12	220	42.8	452	9	AR270635	AR270635 Sequence 6
13	220	42.8	452	6	I49588	I49588 Sequence 6
14	220	42.8	452	9	HUMDFSNSIX	M98331 Homo sapien
15	218.5	42.5	448	9	HUMDEF1A	M21130 Human neutr
16	218.5	42.5	478	6	BD028213	BD028213 Sequence
17	218.5	42.5	482	9	HSHP1	X52053 Human mRNA
18	218.5	42.5	498	9	BD174049	BD174049 Method of
19	218.5	42.5	498	9	HUMDEF1AA	M26602 Human defen
20	218.5	42.5	514	6	AX405718	AX405718 Sequence
21	217.5	42.3	347	9	AF188270	AF188270 Macaca mu
22	216.5	42.1	451	9	HUMDEFB	M23281 Human defen
23	216.5	42.1	464	9	HSDEF3	X13621 Human HNP-3
24	216.5	42.1	556	9	BC027917	BC027917 Homo sapi
25	211.5	41.1	347	9	AF188269	AF188269 Macaca mu
26	207	40.3	424	6	I49587	I49587 Sequence 4
27	202	39.3	485	4	RABDNP4A	M64601 Oryctolagus
28	200	38.9	495	4	RABMCP1A	M28883 Rabbit macr
29	199	38.7	479	4	RABNPCS4A	L10841 Oryctolagus
30	199	38.7	500	4	S55582	S55582 Oryctolagus
31	199	38.7	532	4	RABMCP2A	M28884 Rabbit macr
32	198	38.5	454	4	RABNPCS6A	L10842 Oryctolagus
33	198	38.5	464	4	S55578	S55578 corticostat
34	198	38.5	497	4	RABDNP5A	M64602 Oryctolagus
35	197.5	38.4	543	6	AX588765	AX588765 Sequence
36	196.5	38.2	469	4	RABDNP3AB	M64600 Oryctolagus
37	195	37.9	464	10	RNU16685	U16685 Rattus norv
38	194	37.7	485	10	RNU16686	U16686 Rattus norv
39	186	36.2	504	10	RNU16683	U16683 Rattus norv
40	175.5	34.1	496	9	AF355799	AF355799 Homo sapi
41	173.5	33.8	420	6	AR059558	AR059558 Sequence
42	173.5	33.8	420	6	AR063040	AR063040 Sequence
43	173.5	33.8	420	6	I93525	I93525 Sequence 46
44	171.5	33.4	401	6	AR059566	AR059566 Sequence
45	171.5	33.4	401	6	AR063048	AR063048 Sequence

ALIGNMENTS

RESULT 1

```
A98571
LOCUS       A98571                453 bp    DNA     linear   PAT 26-JAN-2000
DEFINITION  Sequence 2 from Patent WO9911663.
ACCESSION   A98571
VERSION     A98571.1  GI:6781627
KEYWORDS    
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Bougueleret,L. and Chumakov,I.
  TITLE     HUMAN DEFENSIN DEF-X, GENE AND DNAC, COMPOSITION CONTAINING SAME
            AND DIAGNOSTIC AND THERAPEUTIC APPLICATIONS
  JOURNAL   Patent: WO 9911663-A 2 11-MAR-1999;
            BOUGUELERET LYDIE (FR); CHUMAKOV ILYA (FR)
FEATURES             Location/Qualifiers
     source          1..453
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT       107 a     121 c      97 g     128 t
ORIGIN

Alignment Scores:
Pred. No.:                        Length:            453
Score:                 2.7e-55    Matches:           94
Percent Similarity:    100.00%    Conservative:      0
Best Local Similarity: 100.00%    Mismatches:        0
Query Match:           100.00%    Indels:            0
DB:                    6          Gaps:              0

US-10-045-180A-3 (1-94) x A98571 (1-453)

Qy    1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaAlaLeuGlnAlaTrpAlaAlaGlu 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   52 ATGAGGACCCTCACCCTCTCCGCCTTTCTCCTGGTGGCCGCCCTTCAGGCCTGGGCAGAG 111

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  112 CCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGACCAG 171

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  172 GATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGCTCAACA 231

Qy   61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  232 AAGGGCTTGATCTGCCATTGCAGAGTACTATACTGCATTTTGGAGAACATCTTGGTGGG 291

Qy   81 ThrCysPheLeuLeuGlyGluArgTyrProIleCysCysTyr 94
          ||||||||||||||||||||||||||||||||||||
Db  292 ACCTGCTTCATCCTTGGTGAACGCTACCCAATCTGCTGCTAC 333

RESULT 2
BD074746
LOCUS       BD074746              453 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Human defensin polypeptide Def-X, genome DNA and cDNA, composition
            containing the same, and application to diagnosis and remedy
            treatment.
ACCESSION   BD074746
VERSION     BD074746.1  GI:22620349
KEYWORDS    JP 200151264-A/2.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Bougueleret,L. and Shmacov,I.
  TITLE     Human defensin polypeptide Def-X, genome DNA and cDNA, composition
            containing the same, and application to diagnosis and remedy
  JOURNAL   Patent: JP 2001514264-A 2 11-SEP-2001;
            GENSET
COMMENT     OS Homo sapiens (human)
```

```
PN   JP 2001514264-A/2
PD   11-SEP-2001
PF   28-AUG-1998 JP 2000508701
PR   29-AUG-1997 FR   97/10823
PI   LYDIE BOUGUELERET,ILYA SHMACOV
PC   C07K14/435,A01N43/50,A01N63/00,A61K7/00,A61K38/00,
PC   A61P35/00,
PC   A61P37/02,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N15/09, PC
C12P21/02,
PC   C12P21/08,C12Q1/68,G01N33/53,A61K37/02,C12N15/00 CC
Strandedness: Double;
CC   Topology: linear;
CC   Human defensin polypeptide Def-X, genome DNA and cDNA,    CC
     composition
     containing the same, and application to diagnosis and remedy
     treatment
FH   Key             Location/Qualifiers
FT   source          1..453
FT                   /organism='Homo sapiens (human)'.

FEATURES             Location/Qualifiers
     source          1..453
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT       107 a     121 c      97 g     128 t
ORIGIN

Alignment Scores:
Pred. No.:                        Length:            453
Score:                 2.7e-55    Matches:           94
Percent Similarity:    100.00%    Conservative:      0
Best Local Similarity: 100.00%    Mismatches:        0
Query Match:           100.00%    Indels:            0
DB:                    6          Gaps:              0

US-10-045-180A-3 (1-94) x BD074746 (1-453)

Qy    1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaAlaLeuGlnAlaTrpAlaAlaGlu 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   52 ATGAGGACCCTCACCCTCTCCGCCTTTCTCCTGGTGGCCGCCCTTCAGGCCTGGGCAGAG 111

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  112 CCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGACCAG 171

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  172 GATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGCTCAACA 231

Qy   61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  232 AAGGGCTTGATCTGCCATTGCAGAGTACTATACTGCATTTTGGAGAACATCTTGGTGGG 291

Qy   81 ThrCysPheLeuLeuGlyGluArgTyrProIleCysCysTyr 94
          ||||||||||||||||||||||||||||||||||||
Db  292 ACCTGCTTCATCCTTGGTGAACGCTACCCAATCTGCTGCTAC 333

RESULT 3
A98570
LOCUS       A98570               4415 bp    DNA     linear   PAT 26-JAN-2000
DEFINITION  Sequence 1 from Patent WO9911663.
ACCESSION   A98570
VERSION     A98570.1  GI:6781626
KEYWORDS    
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4415)
  AUTHORS   Bougueleret,L. and Chumakov,I.
  TITLE     HUMAN DEFENSIN DEF-X, GENE AND DNAC, COMPOSITION CONTAINING SAME
            AND DIAGNOSTIC AND THERAPEUTIC APPLICATIONS
  JOURNAL   Patent: WO 9911663-A 1 11-MAR-1999;
            BOUGUELERET LYDIE (FR); CHUMAKOV ILYA (FR)
```

```
Qy    65 sHisCysArgValLeuTyrCysIlePheGlyGluHisGlyLeuGlyGlyThrCysPheIleLe 85
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4186 CCATTGCAGAGTACTATACTGCATTTTGGAGAACATCTTGGTGGGACCTGCTTCATCCT 4245

Qy    85 uGlyGluArgTyrProIleCysCysTyr 94
          ||||||||||||||||||||||||||||
Db  4246 TGGTGAACGGCTACCCAATCTGCTGCTAC 4273
```

RESULT 4
BD074745
LOCUS   BD074745   4415 bp   DNA   linear   PAT 27-AUG-2002
DEFINITION   Human defensin polypeptide Def-X, genome DNA and cDNA, composition
             containing the same, and application to diagnosis and remedy
             treatment.
ACCESSION   BD074745
VERSION   BD074745.1   GI:22620348
KEYWORDS   JP 2001514264-A/1.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 4415)
  AUTHORS   Bougueleret,L. and Shmacov,I.
  TITLE   Human defensin polypeptide Def-X, genome DNA and cDNA, composition
             containing the same, and application to diagnosis and remedy
  JOURNAL   Patent: JP 2001514264-A 1 11-SEP-2001;
             GENSET
COMMENT   OS   Homo sapiens (human)
             PN   JP 2001514264-A/1
             PD   11-SEP-2001
             PF   28-AUG-1998 JP 2000508701
             PR   29-AUG-1997 FR   97/10823
             PI   LYDIE BOUGUELERET,ILYA SHMACOV
             PC   C07K14/435,A01N43/50,A01N63/00,A61K7/00,A61K38/00,A61P29/00,
             PC   A61P35/00,
             PC   A61P37/02,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N15/09, PC
             C12P21/02,
             PC   C12P21/08,C12Q1/68,G01N33/53,A61K37/02,C12N15/00 CC
             Strandedness: Double;
             CC   Topology: Linear;
             CC   Human defensin polypeptide Def-X, genome DNA and cDNA,   CC
             composition
             CC   containing the same, and application to diagnosis and remedy
             CC   treatment
             FH   Key   Location/Qualifiers
             FT   exon   1836..1874
             FT   exon   3394..3577
             FT   exon   4161..4380
             FT   CDS   3406..3408
             FT   CDS   4276..4278.

FEATURES   Location/Qualifiers
     source   1..4415
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT   1128 a   1109 c   912 g   1252 t   14 others
ORIGIN

Alignment Scores:
Pred. No.:   5.07e-39   Length: 4415
Score: 393.50   Matches: 94
Percent Similarity: 32.41%   Conservative: 0
Best Local Similarity: 32.41%   Mismatches: 0
Query Match: 76.56%   Indels: 196
DB: Gaps: 1

US-10-045-180A-3 (1-94) x BD074745 (1-4415)

```
Qy     1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaAlaGlnAlaTrpAlaGlu 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3406 ATGAGAGACCCTCACCCTCCTTCTCCTGGTGGCCCTTCAGGCCTGGGCAGAG 3465

Qy    21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProAlaAspAspGln 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3466 CCGCTCCAGCCAAGAGCTCATGAGAGAGTGCCAGCCCAGAGCAGCTCCAGCAGATGACCAG 3525

Qy    41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGln------------- 55
          ||||||||||||||||||||||||||||||||||
Db  3526 GATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGTGAGAGA 3585

Qy    55 ------------------------------------------------------------ 55
Db  3586 TGCCAGCATGCAGAGCTACAGACTAGACAGAGAGCACAGGAGACAGGCTCTGGAATTGGAT 3645

Qy    55 ------------------------------------------------------------ 55
Db  3646 CTCAGTGGCCAGATGTCACTTAGGTGGCTATACTTAACATCTCTGGTCCTGGGATTTTCTCA 3705

Qy    55 ------------------------------------------------------------ 55
Db  3706 TATCTAAATGGAATAGAGAACCAAAGAAATCTAAGAGATTTTCTTTCTCCAAAAACTTG 3765

Qy    55 ------------------------------------------------------------ 55
Db  3766 ATTCCAAGATATGACTGTGAAATTCACTCAGTTAAGATATAAGGAGATGCTACCTAGTT 3825

Qy    55 ------------------------------------------------------------ 55
Db  3826 CCTTTCTGGAGCCAGACAAACAAGCTTAAGTATATAGGAAAATATTTCACCCTGTCTATAT 3885

Qy    55 ------------------------------------------------------------ 55
Db  3886 AGGAGGTTTAGAACCTGGAGGAGGAGCCTAAGAATGTCTTCAGTGTGTGTGATGGGC 3945

Qy    55 ------------------------------------------------------------ 55
Db  3946 AGGAAATGCAGAAAAGTGAAGCAAAGGAATGAGTCTCGAATCCTGTGTGACCAGCACTG 4005

Qy    55 ------------------------------------------------------------ 55
Db  4006 CTCTGTGTATTTATTCCTATTGACTGAGATTGTTTGTGCTACCGGCGTGTAATACAGCCAA 4065

Qy    55 ------------------------------------------------------------ 55
Db  4066 CATCACTCATCAGCCAACATGTGACTTCTCCAAGATTCCCTTTACCACCCACTGCTGNAC 4125

Qy    56 -----------------Val-ProGlySerThrLysGlyLeuIleCy 65
                          ||| |||||||||||||||||||||
Db  4126 CCCGTACTCAGTTTCTGATGCTCTCTCGGGTCCCAGGCTCAACAAAGGGCTTGATCTG 4185
```

FEATURES   Location/Qualifiers
     source   1..4415
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
     exon   1836..1874
     exon   3394..3577
     exon   4161..4380
     polyA_site   4374..4379
BASE COUNT   1128 a   1109 c   912 g   1252 t   14 others
ORIGIN

Alignment Scores:
Pred. No.:   5.07e-39   Length: 4415
Score: 393.50   Matches: 94
Percent Similarity: 32.41%   Conservative: 0
Best Local Similarity: 32.41%   Mismatches: 0
Query Match: 76.56%   Indels: 196
DB: Gaps: 1

US-10-045-180A-3 (1-94) x A98570 (1-4415)

TITLE   Chromosome 8 genomic sequence
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 128544)
AUTHORS   Schudy,A., Schilhabel,M., Schutte,B., Ganz,T., Linzmeier,R.,
          Ho,C.H., Hoang,B.V., McCray,P., Baumgart,A.,
          Schattevoy,R. and Rosenthal,A.
TITLE   Direct Submission
JOURNAL   Submitted (22-FEB-2000) Genome Analysis, Institute of Molecular
          Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE 3 (bases 1 to 128544)
AUTHORS   Genome Sequencing Center Jena.
TITLE   Direct Submission
JOURNAL   Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular
          Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 4 (bases 1 to 128544)
AUTHORS   Genome Sequencing Center Jena.
TITLE   Direct Submission
JOURNAL   Submitted (10-NOV-2000) Genome Analysis, Institute of Molecular
          Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 5 (bases 1 to 128544)
AUTHORS   Lagemann,D. and Platzer,M.
TITLE   Direct Submission
JOURNAL   Submitted (06-JUL-2002) Genome Analysis, Institute of Molecular
          Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 6 (bases 1 to 128544)
AUTHORS   Lagemann,D. and Platzer,M.
TITLE   Direct Submission
JOURNAL   Submitted (02-APR-2003) Genome Analysis, Institute of Molecular
          Biotechnology, Beutenberstr. 11, Jena 07745, Germany
          On Apr 2, 2003 this sequence version replaced gi:21700555.
COMMENT   ---------------------- Genome Center
          Center: Insitute of Molecular Biotechnology
          Center code: IMB
          Web site: http://genome.imb-jena.de/
          Contact: gsci-submit@genome.imb-jena.de
          ---------------------- Project Information
          Center project name: H370
          Center clone name: SCb-561b17
          ---------------------- Summary Statistics
          Sequencing vector: pUC18; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 127999 bases at least Q40
          Consensus quality: 128330 bases at least Q30
          Consensus quality: 128544 bases at least Q20
          Quality coverage: 11.05x

          This sequence was finished as follows unless otherwise noted: all
          regions were double stranded, sequenced with an alternate
          chemistry, or covered by high quality data (i.e. phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest.

          Sequence Quality Assessment:
          This entry has been annotated with sequence quality
          estimates computed by the Phrap assembly program.
          All manually edited bases have been reduced to quality zero.
          Quality levels above 40 are expected to have less than
          1 error in 10,000 bp.
          Base-by-base quality values are not generally visible from the
          GenBank flat file format but are available as part
          of this entry's ASN.1 file.

FEATURES   Location/Qualifiers
     source   1..128544
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="p22-p21"
                     /clone="SCb-561b17"

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3466 CCGCTCCAGCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGACCAG 3525

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGln---------------- 55
                |||||||||||||||||||||||||||||||||||||||||||
Db   3526 GATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCCAGGTTCCAGGTGAGAGA 3585

Qy   55 -- 55

Db   3586 TGCCAGCCATGCAAGCTACAGACTAGACAGAAGGACAGGAGAGACAGGCGCTCGGAATTGGAT 3645

Qy   55 -- 55

Db   3646 CTCAGTGGCAGATGTCACTTAGGTGGCTATACTTAACATCTCTGGTCCTCGGATTTTCTCA 3705

Qy   55 -- 55

Db   3706 TATCTAAATGGAATAGAGAACCAAGAAATCTAAGAGATTTTCTTTTCTCCAAAAACTTG 3765

Qy   55 -- 55

Db   3766 ATTCCAAGATATGACTGTGAAATTCACTAGATTTAAGATATATAAGGAGATGCTACCTAGTT 3825

Qy   55 -- 55

Db   3826 CCTTCTGGAGCCAGACAAACAAGCTTAAGTATATAGGAAAAATATTTCACCCTGTCTATAT 3885

Qy   55 -- 55

Db   3886 AGGAGGTTTTAGAACCTGGAGAGAGCCTAAGAATGTGTTCAGGTGTGTGTGTGATGGGC 3945

Qy   55 -- 55

Db   3946 AGGAATGCAGAAAAGTGAAGCAAAGGAGAGAATGAGTCTCGAATCCTGTGTGACCAGCACTG 4005

Qy   55 -- 55

Db   4006 CTCTGTGTATTTATTCCTATTGACTGAGATTGTTTGTGCTACCGGCTGTAATACAGCCAA 4065

Qy   55 -- 55

Db   4066 CATCACTCATCAGCCAACATGTGACTTCTCCAAGATTCCCTTACCACCCACTGCTGNAC 4125

Qy   56 ------------------------------Val-ProGlySerThrLysGlyLeuIleCy 65
                                                |||||||||||||||||||||||
Db   4126 CCCGTACTCAGTTCTGATGCTCTTCCTGATGTCCCCAGGGTCCACAAGGGCTTGATCTG 4185

Qy   65 sHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIleLe 85
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4186 CCATTGCAGAGTACTATACTGCATTTTGGAGAACATCTTGGTGGGACCTCGTTCATCCT 4245

Qy   85 uGlyGluArgTyrProIleCysCysGlyTyr 94
                |||||||||||||||||||||||||||
Db   4246 TGGTGAACGGCTACCCAATCTGCTGCTAC 4273

RESULT 5
AF238378/c
LOCUS   AF238378   128544 bp   DNA   linear   PRI 02-APR-2003
DEFINITION Homo sapiens chromosome 8 clone SCb-561b17 map p22-p21, complete
            sequence.
ACCESSION   AF238378
VERSION   AF238378.5 GI:29469504
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 128544)
AUTHORS   Polley,A., Baumgart,C., Blechschmidt,K., Dette,M.D., Jahn,N.,
          Menzel,U., Reichwald,K., Schilhabel,M.B., Schudy,A., Taudien,S.,
          Wen,G., Schutte,B., Malik,M., Peng,J.Hong., McCray,P. and
          Rosenthal,A.

```
misc_feature    44529. .44595
                /note="single stranded/single chemistry region"
misc_feature    45703
                /note="low quality region"
misc_feature    45815. .45872
misc_feature    45824. .45872
                /note="low quality region"
                /note="single stranded/single chemistry region"
unsure          45843
unsure          45847
unsure          45849
unsure          45858
unsure          45868
misc_feature    46235
                /note="low quality region"
misc_feature    47899. .47900
                /note="low quality region"
misc_feature    47899. .47900
                /note="low quality region"
misc_feature    49172
                /note="single stranded/single chemistry region"
unsure          49172
misc_feature    49220
                /note="low quality region"
misc_feature    49339
                /note="single stranded/single chemistry region"
misc_feature    49756. .49807
                /note="low quality region"
misc_feature    49961
                /note="single stranded/single chemistry region"
misc_feature    52933. .52948
                /note="low quality region"
unsure          52943
misc_feature    53544
                /note="low quality region"
misc_feature    53612. .53926
                /note="single stranded/single chemistry region"
misc_feature    53674. .53677
                /note="single stranded/single chemistry region"
misc_feature    53779. .53799
                /note="low quality region"
unsure          53787
misc_feature    54734. .55129
                /note="single stranded/single chemistry region"
misc_feature    56028. .56165
                /note="single stranded/single chemistry region"
misc_feature    57952. .58060
                /note="single stranded/single chemistry region"
misc_feature    57968. .57974
                /note="low quality region"
unsure          57968
unsure          57974
misc_feature    58652
                /note="low quality region"
misc_feature    59561. .59638
                /note="single stranded/single chemistry region"
misc_feature    59925
                /note="low quality region"
unsure          59925
misc_feature    59929
misc_feature    61695. .62103
                /note="single stranded/single chemistry region"
misc_feature    61858. .61859
                /note="low quality region"
misc_feature    62263. .62485
                /note="single stranded/single chemistry region"
misc_feature    62655. .63118
                /note="single stranded/single chemistry region"
misc_feature    62750. .62785
                /note="low quality region"
unsure          62774
unsure          62783

misc_feature    62801. .62825
                /note="low quality region"
unsure          62819
unsure          62822
misc_feature    62872. .62874
                /note="low quality region"
misc_feature    62908
                /note="low quality region"
misc_feature    62917. .62987
                /note="low quality region"
unsure          62933. .62935
unsure          62955
unsure          62957
misc_feature    63082
                /note="low quality region"
misc_feature    63110
                /note="low quality region"
misc_feature    63119. .63281
                /note="low quality region"
misc_feature    63119. .63141
                /note="single clone coverage"
unsure          63129. .63131
misc_feature    63142. .63258
                /note="single stranded/single chemistry region"
unsure          63159
unsure          63197. .63258
unsure          63281
misc_feature    63876. .63878
                /note="low quality region"
misc_feature    63933. .64021
                /note="single stranded/single chemistry region"
misc_feature    63943
                /note="low quality region"
misc_feature    63984. .63985
                /note="low quality region"
misc_feature    64004. .64007
                /note="low quality region"
misc_feature    70360
                /note="low quality region"


Alignment Scores:
Pred. No.:                                Length:          128544
Score:                     6.61e-35       Matches:              90
Percent Similarity:        31.60%         Conservative:          1
Best Local Similarity:     31.25%         Mismatches:            2
Query Match:               72.76%         Indels:              196
DB:                                       Gaps:                  1

US-10-045-180A-3 (1-94) x AF238378 (1-128544)

Qy      2 ArgThrThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGluPro 21
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63320 AGGACCCTCACCCTCCTCGCCTTCTCCTGCCTTCTCCGTGGCCCTTCAGCGCTGGGCAGAGCCG 63261

Qy     22 LeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspGlnAsp 41
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db  63260 CTCCCGGCAAGAGCTCATGAGATGCCAGCCCAGAAGAGCAGCCTCCAGCAGATGACCAGGAT 63201

Qy     42 ValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnVal----------- 56
            |||||::|||||||||||||||||||||||||||||||||
Db  63200 GTGGTCCTTTACTTTTCAGGAGATGACAGCTGCTCCTTCAGGT-TCCAGGTGAGAGATG 63142

Qy     56 ------------------------------------------------------ 56

Db  63141 CCAGCATGCAGAGCTTACAGACTAGACAGGAGGACACGGAGCAGGCTCGGAATTGGATCT 63082

Qy     56 ------------------------------------------------------ 56

Db  63081 CAGTGGCAGATGTCACTTAGGTGGCTATACTTAACATCTCTGGTCCTGGATTTTCTCATA 63022

Qy     56 ------------------------------------------------------ 56

Db  63021 TCTAAATGGAATAGAGAACCAAAGAAATCTAAGAGATTTTCTTTCTCCAAAAACTTGAT 62962
```

```
        *   consists of 3 contigs. Gaps between the contigs
        *   are represented as runs of N. The order of the pieces
        *   is believed to be correct as given, however the sizes
        *   of the gaps between them are based on estimates that have
        *   provided by the submittor.
        *   This sequence will be replaced
        *   by the finished sequence as soon as it is available and
        *   the accession number will be preserved.
        *        1    6315: contig of 6315 bp in length
        *     6316    6415: gap of unknown length
        *     6416   16704: contig of 10289 bp in length
        *    16705   16804: gap of unknown length
        *    16805  163093: contig of 146289 bp in length.
FEATURES             Location/Qualifiers
     source          1..163093
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="rp41-339c10"
                     /clone_lib="RPCI - 41 Male (Olive) Baboon BAC Library"
BASE COUNT    44435 a   36508 c   35861 g   46065 t    224 others
ORIGIN

Alignment Scores:
Pred. No.:                                    163093
Score:                  1.74e-25              Length:        163093
Percent Similarity:     300.00                Matches:       79
Best Local Similarity:  28.67%                Conservative:  3
Query Match:            27.62%                Mismatches:    9
DB:                     58.37%                Indels:        195
                        2                     Gaps:          1

US-10-045-180A-3 (1-94) x AC116559 (1-163093)

Qy    4 LeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGluProLeuGln 23
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  32875 CTCACCCCTCCTCTGCCATTCTCCTGGGGCCCTCCCAGGCCTGGACAGAGCCGTCCAG 32934

Qy   24 AlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGlnAspValVal 43
        ||||||||||||||||||||||||||||||||||||::|||||||||
Db  32935 GCAAGAGCTGATGAGATGCCAGCCCCAGGAGCAGCCTCCAGCAAACGCAAACGAAGATGTGGTC 32994

Qy   44 IleTyrPheSerGlyAlaAspAspSerCysSerLeuGlnVal---------- 56
        ||||||||||||||||||||||||||||||||||||||
Db  32995 ATTTACTTTCAGGAGATGACAGCTCCTCTTCCAGGTTCCAGGTGAAAGACACCAGCAT 33054

Qy   56 ----------------------------------------------------- 56
Db  33055 GAAGAGCTACAGACTAGAGGGAAGGGACCGGAGACAGGCTCTGGAATTGGGTCTCAGTGGT 33114

Qy   56 ----------------------------------------------------- 56
Db  33115 ACATGTCACTTAGGTGGCTTACTTACTTAACATCTCTGGTCCTGGTTGTTCTCATGTCTAAAT 33174

Qy   56 ----------------------------------------------------- 56
Db  33175 GGGATAGAGAACCAAATAAATCTAAGAGATTTTCTTCTCCAAAAACATGATTCCAAGA 33234

Qy   56 ----------------------------------------------------- 56
Db  33235 TATGACTGAAATTCACTAGGTTTAAGATATAAGGAGATGCTACCTAGTTCCTTCTGGATC 33294

Qy   56 ----------------------------------------------------- 56
Db  33295 CAGACAAATAAGCTTAAGTATATAGAAAATATTTCACCGTGTCTCATGTAGGAGGTTTTAG 33354

Qy   56 ----------------------------------------------------- 56
Db  33355 AACCTGGAGAGGAGCTAAGAGTGTGTCAAGTGTGTGTGATGGGGTCAGGAGGGGCAAA 33414

Qy   56 ----------------------------------------------------- 56
Db  33415 AAAGTGAGCCAAGGAGAATGAGTTCCGAGTCCTGTGTGACCAGCACTGCTCTGTGTATT 33474
```

```
Qy   56 ----------------------------------------------------- 56
Db  62961 TCCAAGATATGACTGTGAAATTCACTATATTAAGATATAAGGAGATGCTACCTAGTTCC 62902

Qy   56 ----------------------------------------------------- 56
Db  62901 TTCTGGAGCCAGACAAACAAGCTTAAGTATATAGAAAATATTCACCCTGTCTATATAGG 62842

Qy   56 ----------------------------------------------------- 56
Db  62841 AGGTTTTAGAACCTGGAGAGGAGCCTAAGAATGTGTTCAGGTGTGTGTGATGGGCACG 62782

Qy   56 ----------------------------------------------------- 56
Db  62781 AATGACAAAAAGTGAAGCAAAGGAGAATGAGTCTCGAATCCTGTGTGACCAGCACTGCTC 62722

Qy   56 ----------------------------------------------------- 56
Db  62721 TGTGTATTTATTCCTATTGACTGAGATTGTTGTGCTACCGGCTGTAATACAGCCAACAT 62662

Qy   56 ----------------------------------------------------- 56
Db  62661 CACTCATCAGCCAACATGTGACTTCTCCAAGATTCCCTTACCACCCACTGCTGACCCCG 62602

Qy   57 ----------------------ProGlySerThrLysGlyLeuIleCysHis 66
                              ||||||||||||||||||||||||
Db  62601 TACTCAGTTTCTGATGCTCTCTCGGGTCCCCAGGCTCAACAAAGGGCTTGATCTGCCAT 62542

Qy   67 CysArgValLeuTyrCysIlePheGlyIleGluHisLeuGlyGlyThrCysPheLeuLeuGly 86
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62541 TGCAGATGACTATACTGCATTTTTGGAGAACATCTTGGTGGGACCTGCTTCATCCTTGGT 62482

Qy   87 GluArgTyrProIleCysCysTyr 94
        ||||||||||||||||||||||||
Db  62481 GAACGGCTACCCAATCTGCTGCTAC 62458

RESULT 6
AC116559
LOCUS       AC116559      163093 bp    DNA      linear    HTG 30-MAY-2003
DEFINITION  Papio anubis clone rp41-339c10, WORKING DRAFT SEQUENCE, 3 ordered
            pieces.
ACCESSION   AC116559
VERSION     AC116559.18  GI:31193968
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Papio anubis (olive baboon)
ORGANISM    Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 163093)
AUTHORS     Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE       Papio anubis BAC Clone rp41-339c10
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 163093)
AUTHORS     Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (29-MAR-2002) Department Of Chemistry And Biochemistry,
            The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 163093)
AUTHORS     Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (30-MAY-2003) Department of Chemistry And Biochemistry,
            The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 30, 2003 this sequence version replaced gi:30725962.
            --------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR

        * NOTE: This is a 'working draft' sequence. It currently
```

```
BASE COUNT
ORIGIN          /clone lib="RPCI - 41 Male (Olive) Baboon BAC Library"
                53384 a  41769 c  40796 g  52555 t   100 others

Alignment Scores:
Pred. No.:                                      Length:          188604
Score:             2.05e-25                     Matches:             79
Percent Similarity: 300.00                      Conservative:         3
Best Local Similarity: 28.67%                   Mismatches:           9
Query Match:        27.62%                       Indels:             195
DB:                 58.37%                       Gaps:                 1
                    2

US-10-045-180A-3 (1-94) x AC116558 (1-188604)

Qy      4 LeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGluProLeuGln 23
          |||||||||||||||||||||||||||  |||||||||  ||||||
Db 142657 CTCACCCCTCCTCTGCCATTCTCCTGGGGCGGCCCTCCAGGCCTGGACGAGCGCCTCCAG 142598

Qy     24 AlaArgAlaHisGluMetProAlaGlnLysGlnProAlaAspGlnAspValVal 43
          |||||||||||||||||||||::||||::||||||
Db 142597 GCAAGAGCTGATGAGATGCCAGCCCAGGAGCAGCCCCAGCAAACGACCAAGATGTGGTC 142538

Qy     44 IleTyrPheSerGlyAspAspSerCysSerLeuGlnVal------
          ||||||||||||
Db 142537 ATTTACTTTTCAGGAGATGACAGCTCCTCTTCAGGTTCCAGGTGAAAGACACCAGCAT 142478

Qy     56 ----------
Db 142477 GAAGAGCTACAGACTAGAGGGAAGGAGGACGGGAGACAGGCTCGGAATTGGGTCTCAGTGGT 142418

Qy     56 -----------
Db 142417 ACATGTCACTTAGGTGGCTTTACTTAACATCTCTGGTCCTGGTTGTTCTCATGTCTAAAT 142358

Qy     56 ------------
Db 142357 GGGATAGAGAACCAAATAAATCTAAGAGATTTTTCTTTCTCCAAAAAACATGATTCCAAGA 142298

Qy     56 ------------
Db 142297 TATGACTGAAATTCACTAGGTTTAAGATATAAGAGAAGATGCTACCTAGTTCCTTCTGGATC 142238

Qy     56 ------------
Db 142237 CAGACAAATAAGCTTAAGTATATAGAAAATATTCACCGGTGTCTATGTAGGAGGTTTAG 142178

Qy     56 ------------
Db 142177 AACCTGGAGGAGCCTAAGAGTGTGTTCAAGTGTGTGTGATGGGGCAGGAGGAGGCGCAA 142118

Qy     56 ------------
Db 142117 AAAGTGAAGCCAAGGAGGAATGAGTCTCGAGTCCTGTGTGACCAGCACTGCTCTGTGTATT 142058

Qy     56 ------------
Db 142057 TATTCCTATTGACTGAGATTGTTTCTGCTACGGCTGCAATACAGCCAACATCACTCATC 141998

Qy     56 ------------
Db 141997 AGTCAACAGGTGACTTCTCCAAGATTCCCTTACCATCCACTGCGACCCCATACTGAAT 141938

Qy     57 ----------ProGlySerThrLysGlyHisGlyIleLeuIleCysHisCysArgVa 69
          |||||||||||||||||||||||||
Db 141937 TTCTGATGCTCTCTCGTCCCCAGGCTCAACGAAGGGCTTGAGCTGCCATTGCAGAGT 141878

Qy     69 lLeuTyrCysIlePheGlyGluHisLeuGlyThrCysPheLeuGlyGluArgTy 89
          |||||||||||  |||||||||||
Db 141877 ACTCTACTGCCTTTTGGAGAACATCTTGGTGTGGGGACCTCCTTCATCCATGGTGAACGCTC 141818

Qy     89 rProIleCysTyr 94
          ||||||||||
Db 141817 CCCAATCTGCTGCTAC 141802
```

```
Qy     56 ------------
Db  33475 TATTCCTATTGACTGAGATTGTTTCTGCTACCGGCTGCAATACAGCCAACATCACTCATC 33534

Qy     56 ------------
Db  33535 AGTCAACAGGTGACTTCTCCAAGATTCCCTTTACCATCCACTGCGGACCCCATACTGAAT 33594

Qy     57 ----------ProGlySerThrLysGlyHisGlyThrLeuIleCysHisCysArgVa 69
          |||||||||||||||||||||||||
Db  33595 TTCTGATGCTCTCTCGTCCCCAGGCTCAACGAAGGGCTTGAGCTGCCATTGCAGAGT 33654

Qy     69 lLeuTyrCysIlePheGluGlyHisLeuGlyThrCysPheLeuLeuGlyGluArgTy 89
          ||||||  :|||||||||
Db  33655 ACTCTACTGCCTTTTGGAGAACATCTTGGTGTGGGGACCTCCTTCATCCATGGTGAACGCTC 33714

Qy     89 rProIleCysTyr 94
          ||||||||||
Db  33715 CCCAATCTGCTGCTAC 33730
```

```
RESULT 7
AC116558/c
LOCUS       AC116558      188604 bp    DNA     linear   HTG 11-JUN-2003
DEFINITION  Papio anubis clone rp41-273g19, WORKING DRAFT SEQUENCE, 2 ordered
            pieces.
ACCESSION   AC116558
VERSION     AC116558.16  GI:30725961
KEYWORDS    HTG; HTGS_PHASE2; HTGS DRAFT.
SOURCE      Papio anubis (olive baboon)
ORGANISM    Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 188604)
AUTHORS     Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE       Papio anubis BAC Clone rp41-273g19
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 188604)
AUTHORS     Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (29-MAR-2002) Department of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 188604)
AUTHORS     Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (11-JUN-2003) Department of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 15, 2003 this sequence version replaced gi:28173120.
            ------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 2 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *      1     6057: contig of 6057 bp in length
          *   6058     6157: gap of unknown length
          *   6158   188604: contig of 182447 bp in length.
FEATURES             Location/Qualifiers
     source          1..188604
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="rp41-273g19"
```

```
RESULT 8
HSCCSHP4
LOCUS       HSCCSHP4              542 bp    mRNA    linear   PRI 07-JUN-1999
DEFINITION  H.sapiens mRNA for corticostatin HP-4 precursor.
ACCESSION   X65977
VERSION     X65977.1  GI:29734
KEYWORDS    corticostatin/defensin.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 542)
  AUTHORS   Palfree,R.G., Sadro,L.C. and Solomon,S.
  TITLE     The gene encoding the human corticostatin HP-4 precursor contains a
            recent 86-base duplication and is located on chromosome 8
  JOURNAL   Mol. Endocrinol. 7 (2), 199-205 (1993)
  MEDLINE   93225961
   PUBMED   8469233
REFERENCE   2  (bases 1 to 542)
  AUTHORS   Palfree,R.G.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1992) R.G.E. Palfree, McGill University Royal
            Victoria Hosp., Rm L2.05, 687 Pine Ave West, Montreal PQ, H3A 1A1,
            CANADA
FEATURES             Location/Qualifiers
     source          1..542
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="HP-4.1"
                     /tissue_type="bone marrow"
                     /clone_lib="Clontech HL1058a, human bone marrow cDNA"
                     /dev_stage="adult"
     gene            1..542
                     /gene="HP-4"
     CDS             52..345
                     /gene="HP-4"
                     /codon_start=1
                     /product="corticostatin/defensin HP-4 precursor"
                     /protein_id="CAA46792.1"
                     /db_xref="GI:29735"
                     /db_xref="SWISS-PROT:P12838"
                     /translation="MRIIALLAAILLVALQVRAGPLQARGDEAPGQEQRGPEDQDISI
                     SFAWDKSSALQVGSSTRGMVCSCRLVFCRRTELRVGNCLIGGVSFTYCCTRVD"
     sig_peptide     52..108
                     /gene="HP-4"
     mat_peptide     109..342
                     /gene="HP-4"
                     /product="corticostatin/defensin HP-4"
     misc_feature    330..412
                     /gene="HP-4"
                     /note="recent duplication of immediate upstream segment"
BASE COUNT      118 a    137 c    135 g    152 t
ORIGIN

Alignment Scores:
Pred. No.:              3.99e-19        Length:          542
Score:                  227.00          Matches:         46
Percent Similarity:     64.52%          Conservative:    14
Best Local Similarity:  49.46%          Mismatches:      33
Query Match:            44.16%          Indels:          0
DB:                     9               Gaps:            0

US-10-045-180A-3 (1-94) x HSCCSHP4 (1-542)

Qy    1 MetArgThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
        |||||||||   :||  :::||  ||||||||||||||||
Db   52 ATGAGGATTATCGCCCTCCTCGCTGCTATTCTCTTGGTAGCCCTCCAGGTCCGGGCAGGC  111

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspGln  40
        |||  |||  |||  |||:::::  |||
Db      ...
```

```
Db  112 CCACTCCAGGCAAGAGGTGATGAGGCTCCAGGCCAGGAGCAGGCTGGGCAGGAGAAGACCAG  171
Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr  60
        |||  |||  |||||||       |||   :::|||||
Db  172 GACATATCATTTCCTTGCATGGATAAAAGCTCGTTCTTCAGTTTCAGCTCAACA  231

Qy   61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly  80
        :::||||:::||  |||
Db  232 AGGGGCATGGTCTGCTCTGCAGATTAGTATTCTGCCGGCGAACAGAACTTCGTGTTGGG  291

Qy   81 ThrCysPheLeuLeuGlyGluArgTyrProIleCysCys  93
        ||||  :::
Db  292 AACTGCCTCATTGGTGGTGAGTTTCACATACTGCTGC  330

RESULT 9
AF188268
LOCUS       AF188268             347 bp    mRNA    linear   PRI 07-NOV-1999
DEFINITION  Macaca mulatta defensin-1 mRNA, complete cds.
ACCESSION   AF188268
VERSION     AF188268.1  GI:6273105
KEYWORDS    
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 347)
  AUTHORS   Tang,Y.Q., Yuan,J., Miller,C.J. and Selsted,M.E.
  TITLE     Isolation, characterization, cDNA cloning, and antimicrobial
            properties of two distinct subfamilies of alpha-defensins from
            rhesus macaque leukocytes
  JOURNAL   Infect. Immun. 67 (11), 6139-6144 (1999)
  MEDLINE   20002603
   PUBMED   10531277
REFERENCE   2  (bases 1 to 347)
  AUTHORS   Selsted,M.E.
  TITLE     Rhesus myeloid defensins
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 347)
  AUTHORS   Selsted,M.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-1999) Pathology, UC Irvine, College of Medicine,
            Irvine, CA 92697, USA
FEATURES             Location/Qualifiers
     source          1..347
                     /organism="Macaca mulatta"
                     /mol_type="mRNA"
                     /db_xref="taxon:9544"
                     /tissue_type="myeloid"
     CDS             52..342
                     /codon_start=1
                     /product="defensin-1"
                     /protein_id="AAF06312.1"
                     /db_xref="GI:6273106"
                     /translation="MRTLVILAAILLVALQAQAEPLQARTDEATAAQEQIPTDNEPVV
                     VSLAWDESLAPKDSVPGLRKNMACYCRIPACLAGERRYGTCFYLGRVWAFCC"
BASE COUNT       76 a    102 c     94 g     75 t
ORIGIN

Alignment Scores:
Pred. No.:              3.71e-19        Length:          347
Score:                  225.50          Matches:         48
Percent Similarity:     65.62%          Conservative:    15
Best Local Similarity:  50.00%          Mismatches:      30
Query Match:            43.87%          Indels:          3
DB:                     9               Gaps:            2

US-10-045-180A-3 (1-94) x AF188268 (1-347)

Qy    1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
        |||||||||   :|||||  :::||  ||||||||||||||
Db   52 ATGAGGACCCTCGTCATTCTTGCTGCCATTCTCCTGGTGGCCCTCCAGGCGCCTGAG  111
```

RESULT 10

```
Qy    21  ProLeuGlnAlaArgAlaHisGlu---MetProAlaGlnLysGlnProProAlaAspAsp   39
          ||||||||||||:::            |||:::||||:::||||
Db   112  CCACTCCAGGCAAGAACTGACGAGGCTACTGCAGCCCAGGAGCAGATTCCAACAGACAAC  171

Qy    40  GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGln-----ValPro   57
          :::|||||||||:::            |||||||||||
Db   172  CCAGAAGTGGTTGTTTCCCTGCCTGGGATGAAAGCTTGGCTCCAAAGGATTCAGTCCCA  231

Qy    58  GlySerThrLysGlyLeuIleLeuCysHisCysArgValLeuTyrCysIlePheGlyHis   77
          |||                 :::  |||:::||||||
Db   232  GGCTTAAGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCTTAGCAGGAGAACGT  291

Qy    78  LeuGlyGlyThrCysPheLeuGlyGluArgTyrProIleCysCys   93
          |||                     :::
Db   292  CGCTATGGAACCTGCTTCTACCTGGGAAGAGTCTGGCCATTCTGCTGC  339
```

```
RESULT 10
AF184160
LOCUS       AF184160                435 bp    mRNA    linear   PRI 14-NOV-1999
DEFINITION  Macaca mulatta alpha-defensin 1A (MNP1A) mRNA, complete cds.
ACCESSION   AF184160
VERSION     AF184160.1  GI:6409289
KEYWORDS    .
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   Zhao,C., Nguyen,T. and Lehrer,R.I.
  TITLE     cDNA cloning of three alpha-defensins and three demidefensins from
            rhesus monkey bone marrow
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 435)
  AUTHORS   Zhao,C., Nguyen,T. and Lehrer,R.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los
            Angeles, CA 90095, USA
FEATURES             Location/Qualifiers
     source          1..435
                     /organism="Macaca mulatta"
                     /mol_type="mRNA"
                     /db_xref="taxon:9544"
     gene            1..435
                     /gene="MNP1A"
     CDS             13..303
                     /gene="MNP1A"
                     /codon_start=1
                     /product="alpha-defensin 1A"
                     /protein_id="AAF07926.1"
                     /db_xref="GI:6409290"
                     /translation="MRTLAILAAILLVALQAQAEPLQARTDEATAAQEQIPTDNPEVV
                     VSLAWDESLAPKDSVPGLRKNMACYCRIPACLAGERRYGTCFYLGRVWAFCC"

BASE COUNT     120 a   113 c   104 g    98 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                 225.50           Length:        435
Percent Similarity:    65.62%           Matches:        48
Best Local Similarity: 50.00%           Conservative:   15
Query Match:           43.87%           Mismatches:     30
DB:                                     Indels:          3
                                        Gaps:            2

US-10-045-180A-3 (1-94) x AF184160 (1-435)

Qy     1  MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaAlaLeuGlnAlaThrAlaGlu  20
          |||||||||||:::            ::::||||||:::
Db    13  ATGAGGACCCTGACCCTGCTGCCATCCTTGCTGCTGGTGGCCGTCCTGCAGGCCCAGGCTGAG  72

Qy    21  ProLeuGlnAlaArgAlaHisGlu---MetProAlaGlnLysGlnProProAlaAspAsp   39
          ||||||||:::            :::|||||||
Db    73  CCACTCCAGGCAAGAACTGACGAGGCTACTGCAGCCCAGGAGCAGATTCCAACAGACAAC  132

Qy    40  GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGln-----ValPro   57
          :::|||||||||:::      :::
Db    73  CCACTCCAGGCAAGAACTGACGAGGCTACTGCAGCCCAGGAGCAGATTCCAACAGACAAC  132
```

RESULT 11

```
Qy    40  GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGln-----ValPro   57
          :::|||||||||:::            ::::||||:::|||
Db   133  CCAGAAGTGGTTGTTTCCCTGCCTGGGATGAAAGCTTGGCTCCAAAGGATTCAGTCCCA  192

Qy    58  GlySerThrLysGlyLeuIleLeuCysHisCysArgValLeuTyrCysIlePheGlyHis   77
          |||          :::  |||:::||||||
Db   193  GGCTTAAGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCTTAGCAGGAGAACGT  252

Qy    78  LeuGlyGlyThrCysPheLeuGlyGluArgTyrProIleCysCys   93
          |||                     :::
Db   253  CGCTATGGAACCTGCTTCTACCTGGGAAGAGTCTGGCCATTCTGCTGC  300
```

```
RESULT 11
AF184159
LOCUS       AF184159                433 bp    mRNA    linear   PRI 14-NOV-1999
DEFINITION  Macaca mulatta alpha-defensin 1 (MNP1) mRNA, complete cds.
ACCESSION   AF184159
VERSION     AF184159.1  GI:6409287
KEYWORDS    .
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 433)
  AUTHORS   Zhao,C., Nguyen,T. and Lehrer,R.I.
  TITLE     cDNA cloning of three alpha-defensins and three demidefensins from
            rhesus monkey bone marrow
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 433)
  AUTHORS   Zhao,C., Nguyen,T. and Lehrer,R.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los
            Angeles, CA 90095, USA
FEATURES             Location/Qualifiers
     source          1..433
                     /organism="Macaca mulatta"
                     /mol_type="mRNA"
                     /db_xref="taxon:9544"
     gene            1..433
                     /gene="MNP1"
     CDS             13..303
                     /gene="MNP1"
                     /codon_start=1
                     /product="alpha-defensin 1"
                     /protein_id="AAF07925.1"
                     /db_xref="GI:6409288"
                     /translation="MRTLAILAAILLVALQAQAEPLQARTDEATAAQEQIPTDNPEVV
                     VSLAWDESLAPKDSVPGLRKNMACYCRIPACLAGERRYGTCFYMGRVWAFCC"

BASE COUNT     121 a   113 c   102 g    97 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                 223.50           Length:        433
Percent Similarity:    65.62%           Matches:        47
Best Local Similarity: 48.96%           Conservative:   16
Query Match:           43.48%           Mismatches:     30
DB:                                     Indels:          3
                                        Gaps:            2

US-10-045-180A-3 (1-94) x AF184159 (1-433)

Qy     1  MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaAlaLeuGlnAlaThrAlaGlu  20
          |||||||||:::            ::::||||||:::
Db    13  ATGAGGACCCTGACCCTGCTGCCATCCTTGCTGCTGGTGGCCGTCCTGCAGGCCCAGGCTGAG  72

Qy    21  ProLeuGlnAlaArgAlaHisGlu---MetProAlaGlnLysGlnProProAlaAspAsp   39
          ||||||||:::            :::|||||||
Db    73  CCACTCCAGGCAAGAACTGACGAGGCTACTGCAGCCCAGGAGCAGATTCCAACAGACAAC  132

Qy    40  GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGln-----ValPro   57
          :::|||||||||:::      :::
```

```
Db    133  CCAGAAGTGGTTGTTCCCTTGCCTGGATGAAAGCTTGGCTCCAAAGGATTCAGTCCCA 192
Qy     58  GlySerThrLysGlyLeuIleCysHisArgValLeuTyrCysIlePheGlyGluHis 77
                 |||   ||| :::  ||||:::|||   ||||  |||||
Db    193  GGCTTAAGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCTTAGCAGGAGAACGT 252

Qy     78  LeuGlyGlyTyrProIleLeuGlyGluArgTyrProIleCysCys 93
                 |||||||||    :: |||    |||
Db    253  CGCTATGGAACCTGCTTCTACATGGGAAGAGTCTGGCCATTCTGCTGC 300

RESULT 12
AR270635
LOCUS       AR270635              452 bp   DNA      linear   PAT 10-APR-2003
DEFINITION  Sequence 1198 from patent US 6500938.
ACCESSION   AR270635
VERSION     AR270635.1  GI:29701869
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Au-Young,J. and Seilhamer,J.J.
  TITLE     Composition for the detection of signaling pathway gene expression
  JOURNAL   Patent: US 6500938-A 1198 31-DEC-2002;
FEATURES             Location/Qualifiers
     source          1..452
                     /organism="unknown"
BASE COUNT    124 a    118 c    98 g    112 t
ORIGIN

Alignment Scores:
  Pred. No.:               2.46e-18        Length:            452
  Score:                   220.00          Matches:           49
  Percent Similarity:      63.00%          Conservative:      14
  Best Local Similarity:   49.00%          Mismatches:        29
  Query Match:             42.80%          Indels:            8
  DB:                      6               Gaps:              2

US-10-045-180A-3 (1-94) x AR270635 (1-452)

Qy      1  MetArgThrLeuThrLeuThrLeuValLeuValLeuAlaLeuGlnAlaTrpAlaGlu 20
                 ||||||||||||||||||:|||||::||||||||:||
Db     19  ATGAGAACCCTCACCATCCTCACTGCTGTTCTCCTCGTGGCCCTCCAGGCCAAGGCTGAG 78

Qy     21  ----------ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLys 33
                           |||||||:::|||:||||||  ||||||
Db     79  CCACTCCAAGCTGAGGATGATCCACTGCAGGCAAAAGCTTATGAGGCTGATGCCCAGGAG 138

Qy     34  GlnProProAlaAspAspGlnAsnProValValIleTyrPheSerGlyAspAspSerCysSer 53
                 |||  :::  ||||||||
Db    139  CAGCCTGGGGCCAAATGACCAGGACTTTGCCGTCCTTTGCAGAGGATGCAAGCTCAAGT 198

Qy     54  LeuGlnValProGlyServThrLysGlyLeuIleCysHisArgValLeuTyrCysIle 73
                 ||:::::|||      :::  ||||||
Db    199  CTTAGAGCTTTGGGTCAACAAGGGCTTTCACTTGCCATTGCAGA---AGGTCCTGTTAT 255

Qy     74  PheGlyGluHisLeuGlyGlyTyrProIleLeuGlyGluArgTyrProIleCysCys 93
                 ||:::||                 :::  ||||||
Db    256  TCAACAGAATATTCCTATGGGACCTGCACTGTCACTGTATTAACCACAGATTCTGCTGC 315

RESULT 13
I49588
LOCUS       I49588                452 bp   DNA      linear   PAT 07-OCT-1997
DEFINITION  Sequence 6 from patent US 5641497.
ACCESSION   I49588
VERSION     I49588.1  GI:2471808
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Bevins,C.L. and Jones,D.E.
  TITLE     Gastrointestinal defensins, cDNA sequences and method for the
```

```
Db    production and use thereof
  JOURNAL   Patent: US 5641497-A 6 24-JUN-1997;
FEATURES             Location/Qualifiers
     source          1..452
                     /organism="unknown"
BASE COUNT    124 a    118 c    98 g    112 t
ORIGIN

Alignment Scores:
  Pred. No.:               2.46e-18        Length:            452
  Score:                   220.00          Matches:           49
  Percent Similarity:      63.00%          Conservative:      14
  Best Local Similarity:   49.00%          Mismatches:        29
  Query Match:             42.80%          Indels:            8
  DB:                      6               Gaps:              2

US-10-045-180A-3 (1-94) x I49588 (1-452)

Qy      1  MetArgThrLeuThrLeuThrLeuSerAlaPheLeuLeuValLeuAlaLeuGlnAlaTrpAlaGlu 20
                 ||||||||||||||||||:|||||::||||||||:||
Db     19  ATGAGAACCCTCACCATCCTCACTGCTGTTCTCCTCGTGGCCCTCCAGGCCAAGGCTGAG 78

Qy     21  ----------ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLys 33
                           |||||||:::|||:||||||  ||||||
Db     79  CCACTCCAAGCTGAGGATGATCCACTGCAGGCAAAAGCTTATGAGGCTGATGCCCAGGAG 138

Qy     34  GlnProProAlaAspAspGlnAsnProValValIleTyrPheSerGlyAspAspSerCysSer 53
                 |||  :::  ||||||||
Db    139  CAGCCTGGGGCCAAATGACCAGGACTTTGCCGTCCTTTGCAGAGGATGCAAGCTCAAGT 198

Qy     54  LeuGlnValProGlySerThrLysGlyLeuIleCysHisArgValLeuTyrCysIle 73
                 ||:::::|||      :::  ||||||
Db    199  CTTAGAGCTTTGGGTCAACAAGGGCTTTCACTTGCCATTGCAGA---AGGTCCTGTTAT 255

Qy     74  PheGlyGluHisLeuGlyGlyTyrProIleLeuGlyGluArgTyrProIleCysCys 93
                 ||:::||                 :::  ||||||
Db    256  TCAACAGAATATTCCTATGGGACCTGCACTGTCACTGTATTAACCACAGATTCTGCTGC 315

RESULT 14
HUMDFSNSIX
LOCUS       HUMDFSNSIX            452 bp   mRNA      linear   PRI 27-APR-1993
DEFINITION  Homo sapiens defensin 6 mRNA, complete cds.
ACCESSION   M98331
VERSION     M98331.1  GI:181546
KEYWORDS    defensin 6.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Jones,D.E. and Bevins,C.L.
  TITLE     Defensin-6 mRNA in human Paneth cells: implications for
            antimicrobial peptides in host defense of the human bowel
  JOURNAL   FEBS Lett. 315 (2), 187-192 (1993)
   MEDLINE  93114459
   PUBMED   8417977
COMMENT     Original source text: Homo sapiens small intestine cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1..452
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="small intestine"
     5'UTR           1..18
     CDS             19..321
                     /note="prepropeptide"
                     /codon_start=1
                     /product="defensin 6"
                     /protein_id="AAB59357.1"
                     /db_xref="GI:181547"
                     /translation="MRTLTLTLTAVLLTAVLLVALQAKAEPLQAEDDPLQAKAYEADAQEBORGA
                     NDQDPAVSFAEDASSSLRALGSTRAFTCHCRRSCYSTEYSYGTCTVMGINHRFCCL"
     3'UTR           319..440
```

variation   390
              /replace="t"
BASE COUNT 124 a   118 c   98 g   112 t
ORIGIN

Alignment Scores:
Pred. No.:   2.46e-18   Length: 452
Score: 220.00   Matches: 49
Percent Similarity: 63.00%   Conservative: 14
Best Local Similarity: 49.00%   Mismatches: 29
Query Match: 42.80%   Indels: 8
DB: Gaps: 2

US-10-045-180A-3 (1-94) x HUMDFSNSIX (1-452)

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
         ||||||||||||||||||||||:::|||||||||||||||||||
Db   19 ATGAGAACCCTCACCATCCTCACTCTGCTGTTCTCCTCGTTGGCCCTCCAAGGCCAAGGCTGAG 78

Qy   21 --------------ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLys 33
                        ||:::|||:::|||  ||||
Db   79 CCACTCCAAGCTGAGGATGATCCACTGCAGGCAAAAGTTATGAGGCTGAGTGATGCCCAGGAG 138

Qy   34 GlnProProAlaAspAspAspGlnAspValValIleTyrPheSerGlyAspAspSerCysSer 53
               :::        |||||||||||||:::
Db  139 CAGCGTGGGGCAAATGACCAGGACTTTGCCGTCTCCTTTGCAGAGGATGCAAGTCAAGT 198

Qy   54 LeuGlnValProGlySerThrLysGlyLeuIleCysHisCysSerArgValLeuTyrCysIle 73
               |||||||||||||||||||||||
Db  199 CTTAGAGCTTTGGGCTCAACAGGGCTTCACTTGCCAGA---AGGTCCTGTTAT 255

Qy   74 PheGlyGluHisGlyLeuGlyThrCysPheIleLeuGlyGlyArgTyrProIleCysCys 93
               |||:::       :::::||  :::
Db  256 TCAACAGAATATTCCTATGGGACCTGCACTGTCATGGGTATTAACCACAGATTCTGCTGC 315
```

RESULT 15
HUMDEF1A

LOCUS   HUMDEF1A   448 bp   mRNA   linear   PRI 07-NOV-1994
DEFINITION Human neutrophil peptide (defensin) 1 mRNA, complete cds.
ACCESSION   M21130 M22160
VERSION   M21130.1 GI:181526
KEYWORDS   cytotoxic protein; defensin 1; microcidal protein; neutrophil
            peptide.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 448)
AUTHORS   Daher,K.A., Lehrer,R.I., Ganz,T. and Kronenberg,M.
TITLE   Isolation and characterization of human defensin cDNA clones
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85 (19), 7327-7331 (1988)
MEDLINE   89017191
PUBMED   3174637
COMMENT   Original source text: Human leukocyte (myeloid cell line HL-60),
            cDNA to mRNA, clones HNP1a and HNP1b.
            Draft entry and computer-readable sequence [1] kindly submitted by
            K.Daher, 25-NOV-1988.
            Submission form for [1] kindly provided by M.Kronenberg,
            01-JAN-1989.
FEATURES   Location/Qualifiers
     source   1..448
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /map="8p23-p22"
     gene   1..448
                    /gene="DEF1"
     mRNA   <1..448
                    /gene="DEF1"
                    /product="neutrophil peptide 1 mRNA"
     CDS   51..335
                    /gene="DEF1"

                    /note="neutrophil peptide 1 precursor"
                    /codon_start=1
                    /protein_id="AAA52302.1"
                    /db_xref="GI:181527"
                    /db_xref="GDB:G00-120-745"
                    /translation="MRTLAILAAILLVALQAQAEPLQARADEVAAAPEQIAADIPEVV
                    VSLAWDESLAPKHPGSRKNMACYCRIPACIAGERRYGTCIYQGRLWAFCC"
     sig_peptide   51..107
                    /gene="DEF1"
                    /note="neutrophil peptide 1 signal peptide"
     mat_peptide   243..332
                    /gene="DEF1"
                    /product="neutrophil peptide 1"
BASE COUNT   106 a   127 c   110 g   105 t
ORIGIN   Unreported.

Alignment Scores:
Pred. No.:   3.77e-18   Length: 448
Score: 218.50   Matches: 47
Percent Similarity: 64.89%   Conservative: 14
Best Local Similarity: 50.00%   Mismatches: 32
Query Match: 42.51%   Indels: 1
DB: Gaps: 1

US-10-045-180A-3 (1-94) x HUMDEF1A (1-448)

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
         ||||||||||||||||||||:::||||||||||||||||||
Db   51 ATGAGAACCCTCCGCCATCCTTGCCGCCATTCTCCTGGTGGCCCTGCAGGCTGAG 110

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
         ||:::        ||:::|||      |||||
Db  111 CCACTCCAGGCAAGAGCTGATGAGGTTGCGCACCCCGGAGCAGATTGCAGCGGCGGACATC 170

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
               :::||||||||||       :::
Db  171 CCAGAAGTGGTTGTTCCCTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 230

Qy   60 ThrLysGlyLeuIleCysHisCysSerArgValLeuTyrCysIlePheGlyGluHisGlyGly 79
         |||  :::   ||||||:::||||:::
Db  231 AGGAAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTAT 290

Qy   80 GlyThrCysPheIleLeuGlyLeuGlyArgTyrProIleCysCys 93
               |||   |||:::      |||||
Db  291 GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC 332
```

Search completed: December 17, 2003, 16:36:39
Job time : 2738.5 secs

```
              GenCore version 5.1.6
         Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:        December 17, 2003, 14:50:33 ; Search time 3286.5 Seconds
                                             (without alignments)
                                             695.153 Million cell updates/sec

Title:         US-10-045-180A-3
Perfect score: 514
Sequence:      1 MRTLTLLSAFLLVALQAWAE.........GEHLGGTCFILGERYPICCY 94

Scoring table: BLOSUM62
               Xgapop  10.0 , Xgapext    0.5
               Ygapop  10.0 , Ygapext    0.5
               Fgapop   6.0 , Fgapext    7.0
               Delop    6.0 , Delext     7.0

Searched:      22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters:      45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145024_12540/app_query.fasta_1.860
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=-100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10045180 @CGN_1_1_6100 @runat_17122003_145024_12540 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=0 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :     EST:
               1:  em_estba:*
               2:  em_esthum:*
               3:  em_estin:*
               4:  em_estmu:*
               5:  em_estcv:*
               6:  em_estpl:*
               7:  em_estro:*
               8:  em_htc:*
               9:  gb_est1:*
              10:  gb_est2:*
              11:  gb_htc:*
              12:  gb_est3:*
              13:  gb_est4:*
              14:  gb_est5:*
              15:  em_estfun:*
              16:  em_estom:*
              17:  em_gss_hum:*
              18:  em_gss_inv:*
              19:  em_gss_pln:*
              20:  em_gss_vrt:*
              21:  em_gss_fun:*
              22:  em_gss_mam:*
              23:  em_gss_mus:*
              24:  em_gss_pro:*
              25:  em_gss_rod:*
              26:  em_gss_phg:*
              27:  em_gss_vrl:*
              28:  gb_gss1:*
              29:  gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match %	Length	DB	ID	Description
1		227	44.2	422	13	BX099382	BX099382 BX099382
2		227	44.2	634	13	BU616655	BU616655 UI-H-DF0-
3	c	218.5	42.5	349	12	BI021617	BI021617 CM3-MT034
4		218.5	42.5	370	10	BF893383	BF893383 QV3-MT012
5		218.5	42.5	372	10	BF893378	BF893378 QV3-MT012
6		218.5	42.5	379	14	C01361	C01361 HUMGS000834
7		218.5	42.5	387	12	BI021837	BI021837 RCS-MT025
8	c	218.5	42.5	389	12	BI021710	BI021710 CM3-MT034
9	c	218.5	42.5	396	10	BF893221	BF893221 QV3-MT012
10	c	218.5	42.5	413	9	AW468012	AW468012 he30d11.x
11	c	218.5	42.5	419	10	BF893435	BF893435 QV3-MT012
12	c	218.5	42.5	424	10	BF893232	BF893232 QV3-MT012
13	c	218.5	42.5	425	10	BF893231	BF893231 QV3-MT012
14	c	218.5	42.5	430	9	AI684407	AI684407 wa72b06.x
15	c	218.5	42.5	430	9	AI806882	AI806882 wf36d08.x
16	c	218.5	42.5	433	10	BF893440	BF893440 QV3-MT012
17	c	218.5	42.5	449	10	BF893226	BF893226 QV3-MT012
18	c	218.5	42.5	450	10	BF893220	BF893220 QV3-MT012
19	c	218.5	42.5	465	9	AI312619	AI312619 qp77c06.x
20	c	218.5	42.5	472	10	BF893438	BF893438 QV3-MT012
21	c	218.5	42.5	475	10	BF893227	BF893227 QV3-MT012
22		218.5	42.5	486	9	AV741509	AV741509 AV741509
23	c	218.5	42.5	500	13	BU616031	BU616031 UI-H-DF0-
24	c	218.5	42.5	501	13	BU615878	BU615878 UI-H-DF0-
25	c	218.5	42.5	503	12	BM989070	BM989070 UI-H-DF0-
26		218.5	42.5	503	13	BQ188934	BQ188934 UI-E-EJ1-
27		218.5	42.5	505	13	BQ182900	BQ182900 UI-H-EU0-
28		218.5	42.5	506	13	BQ446731	BQ446731 UI-H-EU1-
29	c	218.5	42.5	506	14	CA427527	CA427527 UI-H-DF0-
30	c	218.5	42.5	508	13	BU616082	BU616082 UI-H-DF0-
31	c	218.5	42.5	601	9	AI064980	AI064980 HA0805 Hu
32	c	216.5	42.1	349	12	BI021835	BI021835 RCS-MT025
33	c	216.5	42.1	500	12	BM914478	BM914478 AGENCOURT
34	c	216.5	42.1	504	13	BQ183657	BQ183657 UI-H-EU0-
35		216.5	42.1	555	9	AL697992	AL697992 DKFZp686L
36		216	42.0	542	9	AI250799	AI250799 qi36g07.x
37	c	214.5	41.7	414	9	AW467613	AW467613 he22d02.x
38		214.5	41.7	441	10	BF757978	BF757978 QV3-CT055
39	c	214.5	41.7	461	9	AI091666	AI091666 oo19h12.x
40		214.5	41.7	473	9	AW236289	AW236289 xn72d07.x
41	c	214.5	41.7	480	9	AW236257	AW236257 xn71h04.x
42	c	214.5	41.7	506	13	BQ183268	BQ183268 UI-H-EU0-
43	c	213.5	41.5	421	9	AW291977	AW291977 UI-H-BI2-
44	c	213.5	41.5	461	10	BF906299	BF906299 MR3-MT032
45	c	212.5	41.3	411	9	AW467234	AW467234 he07h02.x

ALIGNMENTS

```
RESULT 1
BX099382
LOCUS       BX099382                422 bp    mRNA    linear   EST 06-FEB-2003
DEFINITION  BX099382 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGp998H134549 ;
            IMAGE:1858620, mRNA sequence.
ACCESSION   BX099382
VERSION     BX099382.1  GI:27844054
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 422)
```

AUTHORS   Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
           Radelof,U., Schneider,D. and Korn,B.
TITLE   Human UnigeneSet - RZPD3
JOURNAL   Unpublished
COMMENT   Contact: Ina Rolfs
           RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
           Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
           RZPD; IMAGp998H134549.
           RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
           Human UnigeneSet - RZPD3 (RZPDLIB No.972)
           http://www.rzpd.de/CloneCards/cgi-
           bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
           RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
           Heubnerweg 6, D-14059 Berlin, Germany
           Tel: +49 30 32639 101
           Fax: +49 30 32639 111
           www.rzpd.de
           This clone is available royalty-free from RZPD;
           contact RZPD (clone@rzpd.de) for further information. Seq primer:
           M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES   Location/Qualifiers
     source   1..422
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGp998H134549 ; IMAGE:1858620"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /clone_lib="Soares_NhHMPu_S1"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
BASE COUNT   83 a   115 c   115 g   107 t   340488-345479, and 484488-489479." 2 others
ORIGIN

Alignment Scores:
Pred. No.:   1.93e-15   Length: 422
Score: 227.00   Matches: 46
Percent Similarity: 64.52%   Conservative: 14
Best Local Similarity: 49.46%   Mismatches: 33
Query Match: 44.16%   Indels: 0
DB: 13   Gaps: 0

US-10-045-180A-3 (1-94) x BX099382 (1-422)

Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
         |||||| :: ||||||||||||||:||||||:||||||||||||||||||||
Db   70 ATGAGGATTATCGCCCTCCTCGCTATTCTTCTGTAGCCCTCCAGGTCCGGGCAGGC 129

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysProProAlaAspAspGln 40
         |||||||||||||| |||||| ||:|: ||||||||
Db   130 CCACTCCAGGCAAGAGAGGTGATGAGGCTCCAGGCAGCGAGGCAGCGTGGGCCAGAAGACCAG 189

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
         || ::::| |||| || |||||| :: |||||||||||
Db   190 GACATATCATTCTTTGCATGGATAAAAAGCTCTGCTCTTCAGTTTCAGGCTCAACA 249

Qy   61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly 80
         :::||:::|| |||||||||:::::: |||||||
Db   250 AGGGGCATGGTCTGCTTGCAGATTAGTATTCTCGCCGGCAACAGAACTTCGTCGTTGGG 309

Qy   81 ThrCysPheLeuLeuGlyLeuArgTyrProIleCysCys 93
         :||||| :::||:: ||| |:||::::
Db   310 AACTGCGCCTCATTGGTGGTGGTGAGTTCACATACTGCTGC 348

RESULT 2
BU616655/c
LOCUS   BU616655   634 bp   mRNA   linear   EST 23-SEP-2002
DEFINITION UI-H-DF0-ben-n-14-0-UI.s1 NCI_CGAP_DF0 Homo sapiens cDNA clone
            UI-H-DF0-ben-n-14-0-UI 3', mRNA sequence.
ACCESSION   BU616655
VERSION   BU616655.1   GI:23282870
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 634)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES   Location/Qualifiers
     source   1..634
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-DF0-ben-n-14-0-UI"
                     /tissue_type="Subchondral Bone"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_DF0"
                     /note="Organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_DF0 is a cDNA library containing the following
                     tissue(s): Subchondral Bone. The library was constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide was used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is GTTAAGCGTC.
                     TAG LIB=UI-H-DF0
                     TAG_TISSUE=subchondral bone
BASE COUNT   170 a   151 c   158 g   155 t   TAG_SEQ=GTTAAGCGTC"
ORIGIN

Alignment Scores:
Pred. No.:   3.38e-15   Length: 634
Score: 227.00   Matches: 46
Percent Similarity: 64.52%   Conservative: 14
Best Local Similarity: 49.46%   Mismatches: 33
Query Match: 44.16%   Indels: 0
DB: 13   Gaps: 0

US-10-045-180A-3 (1-94) x BU616655 (1-634)

Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
         |||||| :: ||||||||||||||:||||||:||||||||||||||||||||
Db   509 ATGAGGATTATCGCCCTCCTCGCTATTCTTCTGTAGCCCTCCAGGTCCGGGCAGGC 450

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysProProAlaAspAspGln 40
         |||||||||||||| |||||: ||:::::|| :::::||
Db   :: :::|||

```
Query Match:      42.51%      Indels:    1
DB:               12          Gaps:      1

US-10-045-180A-3 (1-94) x BI021617 (1-349)

Qy      1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
          |||||||||||||||||| :::||||||||||||||||||||  ||||||||||||||||
Db    311 ATGAGGACCCTGACCCTGCTGCCATCCTTCCTGGTGGCCCTGCAGGCCTGGGCTGAG 252

Qy     21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
          ||||||||| |||||||||| ||||    ::|||  ||| ||||||||||||||||||||
Db    251 CCACTCCAGGCAAGAGCTGATGAGGTTGCGAGCCCCGAGCCAGATTGCAGCGGACATC 192

Qy     40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
          ||||||| ::||||||||||||||||||||  :::|||||||||||||||||||||||||
Db    191 CCAGAAGTGGTTGTTCCCTTGCATGGGACGAAAGCTCGGCTCCAAAGCATCCAGGCTCA 132

Qy     60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
          |||||||||||||||:::|| ||||||||||||||||| :::|||||||||||| |||||
Db    131 AGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGGTGCATTGCAGAGAACGTCGCTAT 72

Qy     80 GlyThrCysPheIleLeuGlyLeuArgTyrProIleCysCys 93
          ||||||||:::||| ||||||||||| ::: |||||||
Db     71 GGAACCTGCATCTACCAGGGAAGACTCTGGGCCATTCTGCTGC 30
```

```
RESULT 4
BF893383
LOCUS         BF893383                 370 bp    mRNA       linear      EST 18-JAN-2001
DEFINITION    QV3-MT0129-111100-427-g01 MT0129 Homo sapiens cDNA, mRNA sequence.
ACCESSION     BF893383
VERSION       BF893383.1  GI:12284842
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 370)
  AUTHORS     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
              Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
              ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.
  TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
  JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE     20202663
  PUBMED      10737800
  COMMENT     Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2707001
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPESP/LICR Human Cancer Genome
              Project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3&t2=QV3-MT0129-
              111100-427-g01&t3=2000-11-11&t4=1)
              Seq primer: puc 18 forward
              High quality sequence stop: 370.
  FEATURES             Location/Qualifiers
       source          1..370
                       /organism="Homo sapiens"
                       /mol_type="mRNA"
                       /db_xref="taxon:9606"
                       /dev_stage="Adult"
                       /clone_lib="MT0129"
                       /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                       SmaI; A mini-library was made by cloning products derived
                       from ORESTES PCR (U.S. Letters Patent application No. 196
                       ,716 - Ludwig Institute for Cancer Research) profiles
```

```
Db    449 CCACTCCAGGCAAGAGGTGATGAGGCTCCAGGCAGCAGCGTGGGCCAGAAGACCAG 390

Qy     41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
          |||:::: |||  |||||||   ::||||||||||||||||||||  ||||||||||
Db    389 GACATARTCTATTTCCTTTGCATGGGATAAAAAGCTCGGCTCTTCCAGGTTTCAGGCTCAACA 330

Qy     61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly 80
          ||:||||||||||| ||:::::||||||||||||||:::||||||||||||| |||
Db    329 AGGGGCATGGTCTGCTCTTGCAGATTAGTATTCTGCCGGCGAACAGAACTTCGTGTTGGG 270

Qy     81 ThrCysPheIleLeuGlyLeuGluArgTyrProIleCysCys 93
          ||||||||||||| |||  ::: |||||||
Db    269 AACTGCCTCATTGGCGGTGTGAGTTTCACATACTGCTGC 231
```

```
RESULT 3
BI021617/c
LOCUS         BI021617                 349 bp    mRNA       linear      EST 14-JUN-2001
DEFINITION    CM3-MT0343-170101-665-a05 MT0343 Homo sapiens cDNA, mRNA sequence.
ACCESSION     BI021617
VERSION       BI021617.1  GI:14428247
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 349)
  AUTHORS     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
              Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
              ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.
  TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
  JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE     20202663
  PUBMED      10737800
  COMMENT     Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2707001
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPESP/LICR Human Cancer Genome
              Project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM3&t2=CM3-MT0343-
              170101-665-a05&t3=2001-01-17&t4=1)
              Seq primer: puc 18 forward
              High quality sequence stop: 349.
  FEATURES             Location/Qualifiers
       source          1..349
                       /organism="Homo sapiens"
                       /mol_type="mRNA"
                       /db_xref="taxon:9606"
                       /dev_stage="Adult"
                       /clone_lib="MT0343"
                       /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                       SmaI; A mini-library was made by cloning products derived
                       from ORESTES PCR (U.S. Letters Patent application No. 196
                       ,716 - Ludwig Institute for Cancer Research) profiles
                       into the pUC 18 vector. Reverse transcription of tissue
                       mRNA and cDNA amplification were performed under low
                       stringency conditions."
BASE COUNT     69 a      98 c      98 g      84 t
ORIGIN

Alignment Scores:
Pred. No.:              1.31e-14          Length:           349
Score:                  218.50            Matches:          47
Percent Similarity:     64.89%            Conservative:     14
Best Local Similarity:  50.00%            Mismatches:       32
```

into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT   80 a   109 c   105 g   76 t
ORIGIN

Alignment Scores:
Pred. No.:   1.42e-14   Length: 370
Score: 218.50   Matches: 47
Percent Similarity: 64.89%   Conservative: 14
Best Local Similarity: 50.00%   Mismatches: 32
Query Match: 42.51%   Indels: 1
DB: Gaps: 1

US-10-045-180A-3 (1-94) x BF893383 (1-370)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
         |||||||||||||||||||| ::||||||||||||||
Db   87 ATGAGGACCCTCGCCATCCTTGCTGCCATTCTCCTGGTCGCCCTGCAGGCCCAGGCTGAG 146

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
         ||||||||||||||||||:::||| :::||||
Db 147 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGAGCAGACGCGGACATC   206

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
         ::||||||||| ::: ||||||| :|| ::: ||||||||||
Db 207 CCAGAAGTGGTTGTTTCCCTTGCATGGGACGAAAAGCTTGGCTCCAAAGCATCCAGGCTCA 266

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
         ||| |||||||||||::: |||||||||:::
Db 267 AGGAAAACATGGCCTGCTATTGCAGAATACCAGCCGTGCATTGCAGGAGAACGTCGCTAT 326

Qy   80 GlyThrCysPheLeuLeuGlyGluArgTyrProIleCysCys 93
         |||||||||||||
Db 327 GGAACCTGCATCTACCAGGGAAGACTCTGGGCCATTCTGCTGC 368

RESULT 5
BF893378
LOCUS   BF893378   372 bp   mRNA   linear   EST 18-JAN-2001
DEFINITION QV3-MT0129-111100-427-c03 MT0129 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF893378
VERSION   BF893378.1 GI:12284837
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 372)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3&t2=QV3-MT0129-
            11100-427-c03&t3=2000-11-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 372.

FEATURES   Location/Qualifiers
     source   1..372
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0129"
                     /note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."

BASE COUNT   80 a   109 c   106 g   77 t
ORIGIN

Alignment Scores:
Pred. No.:   1.43e-14   Length: 372
Score: 218.50   Matches: 47
Percent Similarity: 64.89%   Conservative: 14
Best Local Similarity: 50.00%   Mismatches: 32
Query Match: 42.51%   Indels: 1
DB: Gaps: 1

US-10-045-180A-3 (1-94) x BF893378 (1-372)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
         |||||||||||||||||||| ::||||||||||||||
Db   85 ATGAGGACCCTCGCCATCCTTGCTGCCATTCTCCTGGTCGCCCTGCAGGCCCAGGCTGAG 144

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
         ||||||||||||||||||:::||| :::||||
Db 145 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGAGCAGATTGCAGCGGACATC 204

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
         ::||||||||| ::: ||||||| :|| ::: ||||||||||
Db 205 CCAGAAGTGGTTGTTTCCCTTGCATGGGACGAAAAGCTTGGCTCCAAAGCATCCAGGCTCA 264

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
         ||| |||||||||||::: |||||||||:::
Db 265 AGGAAGAACATGGCCTGCTATTGCAGAATACCAGCCGTGCATTGCAGGAGAACGTCGCTAT 324

Qy   80 GlyThrCysPheLeuLeuGlyGluArgTyrProIleCysCys 93
         |||||||||||||
Db 325 GGAACCTGCATCTACCAGGGAAGACTCTGGGCCATTCTGCTGC 366

RESULT 6
C01361
LOCUS   C01361   379 bp   mRNA   linear   EST 31-DEC-2002
DEFINITION HUMGS0008342 Human adult (K.Okubo) Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   C01361
VERSION   C01361.1 GI:1433591
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 379)
  AUTHORS   Okubo,K.
  TITLE   BodyMap; human gene expression database
  JOURNAL   Unpublished
COMMENT   Contact: Okubo,K.
            Institute for Molecular and Cellular Biol
            Osaka University
            1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
            Tel: 06-877-5111(ex.3315)
            Email: kousaku@imcb.osaka-u.ac.jp
            We are not submitting the same cDNA sequence redundantly to DDBJ
            since 1993. For the abundance information of clones with this
            sequence in this library and as well in other 3'-directed
            libraries, see', http://www.imcb.osaka-u.ac.jp/bodymap'. The

sequences of the clones represented by this GS sequences is also
found there.

```
FEATURES
    source          Location/Qualifiers
                    1..379
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /dev_stage="adult"
                    /clone_lib="Human adult (K.Okubo)"
                    /note="One or more human adult tissue"

BASE COUNT        90 a      102 c      100 g       82 t       5 others
ORIGIN

Alignment Scores:
Pred. No.:                            Length:            379
Score:                 218.50        Matches:            47
Percent Similarity:    64.89%        Conservative:       14
Best Local Similarity: 50.00%        Mismatches:         32
Query Match:           42.51%        Indels:              1
DB:                                  Gaps:                1

US-10-045-180A-3 (1-94) x C01361 (1-379)

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
        |||||||  |||   :||:|||:|||  ||||||||||||||||||||||||||||||
Db   18 ATGAGGACCCTGCCATCCTTGCTGCCATTCTCTGGTGGCCCTGCGCAGGCCCAGGCTGAG  77

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp  39
        ||||||||:|||  |||   ::|:|      |||  |::||||||||||||||||||||
Db   78 CCACTCCAGGCAAGAGCTGATGAGGTTGCTCGCAGCCCCGGAGCAGATTGCAGCGGACATC 137

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer  59
        :::|||||||   |:|||| :|| :::   ||| ::: :|| ||||||
Db  138 CCAGAAGTGGTTGTTCCCTGCATGGGACGAAAGCATCCGGCTCCAAAGCATCCAGGCTCA 197

Qy   60 ThrLysGlyLeuIleCysHisCysSerArgValLeuTyrCysIlePheGlyGluHisLeuGly  79
        |||   :::  |||||||:|  ||||| |||||||
Db  198 AGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTAT 257

Qy   80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys  93
        ||||||||  ::: || |:||||||||||||||
Db  258 GGNACCTGCATCTACCAGGAGAAGACTCTGGGCATTCTGCTGC 299

RESULT 7
BI021837
LOCUS        BI021837              387 bp    mRNA      linear   EST 14-JUN-2001
DEFINITION   RCS-MT0259-170101-011-B02 MT0259 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BI021837
VERSION      BI021837.1  GI:14428467
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 387)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
TITLE        Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
PUBMED       10737800
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
```

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be derived from the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RCS&t2=RCS-MT0259-
170101-011-B02&t3=2001-01-17&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 387.

```
FEATURES
    source          Location/Qualifiers
                    1..387
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /dev_stage="Adult"
                    /clone_lib="MT0259"
                    /note="Organ: marrow; Vector: puc18; Site 1: SmaI, Site 2:
                    SmaI; A mini-library was made by cloning products derived
                    from ORESTES PCR (U.S. Letters Patent application No. 196
                    ,716 - Ludwig Institute for Cancer Research) profiles
                    into  the pUC 18 vector. Reverse transcription of tissue
                    mRNA and cDNA amplification were performed under low
                    stringency conditions."

BASE COUNT        96 a      104 c      104 g       83 t
ORIGIN

Alignment Scores:
Pred. No.:                            Length:            387
Score:                 218.50        Matches:            47
Percent Similarity:    64.89%        Conservative:       14
Best Local Similarity: 50.00%        Mismatches:         32
Query Match:           42.51%        Indels:              1
DB:                                  Gaps:                1

US-10-045-180A-3 (1-94) x BI021837 (1-387)

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
        |||||||  |||   :||:|||:|||  ||||||||||||||||||||||||||||||
Db   20 ATGAGGACCCTGCCATCCTTGCTGCCATTCTCTGGTGGCCCTGCGCAGGCCCAGGCTGAG  79

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp  39
        ||||||||:|||  |||   ::|:|      |||  |::||||||||||||||||||||
Db   80 CCACTCCAGGCAAGAGCTGATGAGGTTGCTCGCAGCCCCGGAGCAGATTGCAGCGGACATC 139

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer  59
        :::|||||||   |:|||| :|| :::   ||| ::: :|| ||||||
Db  140 CCAGAAGTGGTTGTTCCCTGCATGGGACGAAAGCATCCGGCTCCAAAGCATCCAGGCTCA 199

Qy   60 ThrLysGlyLeuIleCysHisCysSerArgValLeuTyrCysIlePheGlyGluHisLeuGly  79
        |||   :::  |||||||:|  ||||| |||||||
Db  200 AGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTAT 259

Qy   80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys  93
        ||||||||  ::: || |:||||||||||||||
Db  260 GGAACCTGCATCTACCAGGAGAAGACTCTGGGCATTCTGCTGC 301

RESULT 8
BI021710/c
LOCUS        BI021710              389 bp    mRNA      linear   EST 14-JUN-2001
DEFINITION   CM3-MT0343-170101-666-F02 MT0343 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BI021710
VERSION      BI021710.1  GI:14428340
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 389)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
TITLE        Shotgun sequencing of the human transcriptome with ORF expressed
```

JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
PUBMED   10737800
COMMENT   sequence tags
             Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=CM3-MT0343-
             170101-666-f02&t3=2001-01-17&t4=1)
             Seq primer: puc 18 forward
             High quality sequence start: 52
             High quality sequence stop: 389.
FEATURES   Location/Qualifiers
     source 1..389
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /dev_stage="Adult"
             /clone_lib="MT0343"
             /note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
             SmaI; A mini-library was made by cloning products derived
             from ORESTES PCR (U.S. Letters Patent application No. 196
             ,716 - Ludwig Institute for Cancer Research) profiles
             into the pUC 18 vector. Reverse transcription of tissue
             mRNA and cDNA amplification were performed under low
             stringency conditions."
BASE COUNT   81 a   106 c   106 g   96 t
ORIGIN

Alignment Scores:
Pred. No.:   389
Score: 1.52e-14   Length: 389
                    218.50   Matches: 47
Percent Similarity: 64.89%   Conservative: 14
Best Local Similarity: 50.00%   Mismatches: 32
Query Match: 42.51%   Indels: 1
DB: Gaps: 12

US-10-045-180A-3 (1-94) x BI021710 (1-389)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGluGlu 20
        | |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 351 ATGAGGACCCTCGCCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCTGAG 292

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
        || ||||||||||||| :::|| :|||||||||||||||
Db 291 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGGAGCAGCGGACATC 232

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
        ::|||||| ::: ||||||| ::: ||||||||
Db 231 CCGAAGTGGTTGTTTCCCTTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 172

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyLeuHisLeuGly 79
        ::|||||||||||||||||||| :::||| ||||||
Db 171 AGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGGAACGTCGCTAT 112

Qy   80 GlyThrLysPheIleLeuGlyGluArgTyrProIleCysCys 93
        :: ||||||||||| :||:
Db 111 GGAACCTGCCATCTACCAGGGAAGACTTCGGGCATTCTGCTGC 70

RESULT 9
BF893221/c
LOCUS   BF893221   396 bp   mRNA   linear   EST 18-JAN-2001
DEFINITION QV3-MT0129-111100-419-d01 MT0129 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF893221
VERSION   BF893221.1   GI:12284680
KEYWORDS   EST.

SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 396)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=QV3-MT0129-
             111100-419-d01&t3=2000-11-11&t4=1)
             Seq primer: puc 18 forward
             High quality sequence start: 38
             High quality sequence stop: 396.
FEATURES   Location/Qualifiers
     source 1..396
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /dev_stage="Adult"
             /clone_lib="MT0129"
             /note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
             SmaI; A mini-library was made by cloning products derived
             from ORESTES PCR (U.S. Letters Patent application No. 196
             ,716 - Ludwig Institute for Cancer Research) profiles
             into the pUC 18 vector. Reverse transcription of tissue
             mRNA and cDNA amplification were performed under low
             stringency conditions."
BASE COUNT   88 a   99 c   111 g   98 t
ORIGIN

Alignment Scores:
Pred. No.:   396
Score: 1.56e-14   Length: 396
                    218.50   Matches: 47
Percent Similarity: 64.89%   Conservative: 14
Best Local Similarity: 50.00%   Mismatches: 32
Query Match: 42.51%   Indels: 1
DB: Gaps: 10

US-10-045-180A-3 (1-94) x BF893221 (1-396)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
        | |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 394 ATGAGGACCCTCGCCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCTGAG 335

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
        || ||||||||||||| :::|| :|||||||||||||||
Db 334 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGGAGCAGCGGACATC 275

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
        ::|||||| ::: ||||||| ::: ||||||||
Db 274 CCGAAGTGGTTGTTTCCCTTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 215

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyLeuHisLeuGly 79
        ::|||||||||||||||||||| :::||| ||||||
Db 214 AGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGGAACGTCGCTAT 155

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
          ||||||||||||||||||||||| ::::||||||||||||| ||||||||
Db   225 AGGAAAACATGGCCTGCTATTGCAGATTACCAGCGTGCATTGCGAGGAAACGTCGCTAT 166

Qy   80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
          ||||||||||||||||||||||||||||||||||||||||||
Db   165 GGAACCTGCATCTACCAGGGAGACTCTGGGCATTCTGCTGC 124

RESULT 11
BF893435/c
LOCUS   BF893435   419 bp   mRNA   linear   EST 18-JAN-2001
DEFINITION   QV3-MT0129-111100-426-c04 MT0129 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF893435
VERSION   BF893435.1 GI:12284894
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 419)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/getthtml2.pl?tl=QV3&t2=MT0129-
             111100-426-c04&t3=2000-11-11&t4=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 419.
FEATURES   Location/Qualifiers
     source   1..419
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MT0129"
                     /dev_stage="Adult"
                     /note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT   101 a   103 c   114 g   101 t
ORIGIN

Alignment Scores:
Pred. No.:   1.69e-14   Length: 419
Score: 218.50   Matches: 47
Percent Similarity: 64.89%   Conservative: 14
Best Local Similarity: 50.00%   Mismatches: 32
Query Match: 42.51%   Indels: 1
DB: 10   Gaps: 1

US-10-045-180A-3 (1-94) x BF893435 (1-419)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20

Qy   80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
          ||||||||||||||||||||||||||||||||||| :::
Db   154 GGAACCTGCATCTACCAGGGAGAGACTCTGGGCATTCTGCTGC 113

RESULT 10
AW468012/c
LOCUS   AW468012   413 bp   mRNA   linear   EST 24-FEB-2000
DEFINITION   he30d11.x1 NCI_CGAP CML1 Homo sapiens cDNA clone IMAGE:2920533 3'
             similar to gb:M26602 NEUTROPHIL DEFENSINS 1, 2 AND 3 PRECURSOR
             (HUMAN), mRNA sequence.
ACCESSION   AW468012
VERSION   AW468012.1 GI:7038118
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 413)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Elisabeth Paietta, Jonathan D. Licht, M.D.
             Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
             Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
             I.M.A.G.E. Consortium cDNA Sequencing by: Washington University
             Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www.bio.llnl.gov/bbrp/image/image.html
             Seq primer: -40UP from Gibco.
FEATURES   Location/Qualifiers
     source   1..413
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2920533"
                     /tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
                     rearrangement positive, includes both chronic phase and
                     myeloid blast crisis"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP CML1"
                     /note="Organ: whole blood; Vector: pCMV-SPORT6; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally. Primer:
                     Oligo dT. Library constructed by Life Technologies."
BASE COUNT   95 a   101 c   113 g   103 t   1 others
ORIGIN

Alignment Scores:
Pred. No.:   1.65e-14   Length: 413
Score: 218.50   Matches: 47
Percent Similarity: 64.89%   Conservative: 14
Best Local Similarity: 50.00%   Mismatches: 32
Query Match: 42.51%   Indels: 1
DB: 9   Gaps: 1

US-10-045-180A-3 (1-94) x AW468012 (1-413)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
          |||||||||||||| :::||||||::::| |||||||||||||||||||||||||
Db   405 ATGAGGACCCTGACCCTGCTGCCATCCTTCCTGTGGCCCTGCCGGGCNCAGGCTGAG 346

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
          |||||||||||||||||| ::::||| |||||||
Db   345 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCCGGAGCAGATTGCAGCGGACATC 286

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
          |||||||| ||||||||| ||||::: ||||| |||
Db   285 CCAGAAGTGGTTGTTTTCCCTTGCATGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 226

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20

```
Alignment Scores:
Pred. No.:                1.72e-14        Length:         424
Score:                    218.50          Matches:        47
Percent Similarity:       64.89%          Conservative:   14
Best Local Similarity:    50.00%          Mismatches:     32
Query Match:              42.51%          Indels:         1
DB:                       10              Gaps:           1

US-10-045-180A-3 (1-94) x BF893232 (1-424)

Qy    1  MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu   20
         ||||||||||||||||  :::||:||||| |||||||||   |||| ||||||||||||
Db  383  ATGAGGACCCTGACCCTGCTGCCATCTTCCTTGCTGGCCCTGGCCCTGCAGGCCTGGGCTGAG  324

Qy   21  ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp   39
         |||||| ||| :: ||||||||||    |||||   ::::|||| |||||||||||||
Db  323  CCACTCCAGGCAAGAGCTGATGAGGTTGCTCCAGCCCCGGAGCAGATTGCAGCGGGACATC  264

Qy   40  GlnAspValValIleTyrPheSerGlyAspSerCysSerLeuGlnValProGlySer   59
         ::: ||| |||||| ::: :::    :::  ||||||||||||||
Db  263  CCAGAAGTGGTTGTTTCCCTGCATGGGACGAAAGCATCCAGGCTCA  204

Qy   60  ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyHisLeuGly   79
         ||| ::::|||||| |||::: :::|||| ||||| |||||
Db  203  AGGAAAAACATGGCCGCTATTGCAGAATACCAGCGGTCATTGCAGGAGAACGTCGCTAT  144

Qy   80  GlyThrCysPheLeuLeuGlyGluArgTyrProIleCysCys   93
         :::  ||||||||||||||||| :::  |||||
Db  143  GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC  102

RESULT 13
BF893231/c
LOCUS       BF893231                  425 bp    mRNA         linear    EST 18-JAN-2001
DEFINITION  QV3-MT0129-111100-419-h03 MT0129 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF893231
VERSION     BF893231.1  GI:12284690
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   MEDLINE  20202663
   PUBMED   10737800
  COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPSP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3&t2=QV3-MT0129-
            111100-419-h03&t3=2000-11-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 23
            High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1..425
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
```

```
Db  413  ATGAGGACCCTGCCATCCTTGCTGCCATCTTCCTTGGTGGCCCTGGCCCTGCAGGCCTGAG  354

Qy   21  ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp   39
         |||||| ||| :: ||||||||||    |||||   ::::|||| |||||||||||||
Db  353  CCACTCCAGGCAAGAGCTGATGAGGTTGCTCCAGCCCCGGAGCAGATTGCAGCGGGACATC  294

Qy   40  GlnAspValValIleTyrPheSerGlyAspSerCysSerLeuGlnValProGlySer   59
         ::: ||| |||||| ::: :::    :::  ||||||||||||||
Db  293  CCAGAAGTGGTTGTTTCCCTGCATGGGACGAAAGCATCCAGGCTCA  234

Qy   60  ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyHisLeuGly   79
         ||| :: ||||||| |||::: :::|||| ||||| |||||
Db  233  AGGAAAAACATGGCCGCTATTGCAGAATACCAGCGGTCATTGCAGGAGAACGTCGCTAT  174

Qy   80  GlyThrCysPheLeuLeuGlyGluArgTyrProIleCysCys   93
         :::  ||||||||||||||||| :::  |||||
Db  173  GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC  132

RESULT 12
BF893232/c
LOCUS       BF893232                  424 bp    mRNA         linear    EST 18-JAN-2001
DEFINITION  QV3-MT0129-111100-419-h09 MT0129 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF893232
VERSION     BF893232.1  GI:12284691
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   MEDLINE  20202663
   PUBMED   10737800
  COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPSP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3&t2=QV3-MT0129-
            111100-419-h09&t3=2000-11-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 21
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1..424
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MT0129"
                     /dev_stage="Adult"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      96 a    107 c   119 g   102 t
ORIGIN
```

```
                              /db_xref="taxon:9606"
                              /dev_stage="Adult"
                              /clone_lib="MTO129"
                              /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                              SmaI; A mini-library was made by cloning products derived
                              from ORESTES PCR (U.S. Letters Patent application No. 196
                              ,716 - Ludwig Institute for Cancer Research) profiles
                              into the pUC 18 vector. Reverse transcription of tissue
                              mRNA and cDNA amplification were performed under low
                              stringency conditions."
BASE COUNT       96 a       108 c       118 g       103 t
ORIGIN

Alignment Scores:
Pred. No.:               1.72e-14      Length:            425
Score:                   218.50        Matches:           47
Percent Similarity:      64.89%        Conservative:      14
Best Local Similarity:   50.00%        Mismatches:        32
Query Match:             42.51%        Indels:            1
DB:                      10            Gaps:              1

US-10-045-180A-3 (1-94) x BF893231 (1-425)

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
         ||||||||||||||||   :::|||||  |||||||||||||||||||||||||||||||
Db  387 ATGAGGACCCTCGCCATCCTTGCCATCTTCCCTGCGTGGGCCCTGCAGGCCTGAG 328

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp  39
         ||||||||||||  |||:::   ::|||     ||:::|||||||||||||||||||||
Db  327 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGGAGCAGATTGCAGCGGACATC 268

Qy   40 GlnAspValValIleTyrPheSerGlyAspSerCysSerLeuGlnValProGlySer  59
         ||||||||||||:::  |||  |||  :::  |||||||||||||||||||||||||
Db  267 CCAGGAGTGGTTGTTTCCCTGCGGACGAGCTTGGCTCCAAAGCATCCAGGCTCA 208

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGluGlyHisLeuGly  79
         ||| :::||||||||||:::    ::: |||||||||  |||||||
Db  207 AGGAAAAACATGCCCTGCTATTACCAGCGTGCATTCTGCTGC 148

Qy   80 GlyThrCysPheIleLeuGlyLeuGluArgTyrProIleCysCys  93
         |||||   |||||||   :::   ||||||||||||||||
Db  147 GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC 106

RESULT 14
AI684407/c
LOCUS     AI684407            430 bp    mRNA      linear    EST 07-MAR-2000
DEFINITION  wa72b06.x1 Soares NFL T GBC S1 Homo sapiens cDNA clone
            IMAGE:2301683 3' similar to gb:M26602 NEUTROPHIL DEFENSINS 1, 2 AND
            3 PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   AI684407
VERSION     AI684407.1      GI:4895701
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 430)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 526   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1..430
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2301683"
```

```
                              /lab_host="DH10B"
                              /clone_lib="Soares NFL T GBC S1"
                              /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                              a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                              Equal amounts of plasmid DNA from three normalized
                              libraries (fetal lung NbHL19W, testis NHT, and B-cell
                              NCI_CGAP_GCB1) were mixed, and 88 circles were made in
                              vitro. Following HAP purification, this DNA was used as
                              tracer in a subtractive hybridization reaction. The driver
                              was PCR-amplified cDNAs from pools of 5,000 clones made
                              from the same 3 libraries. The pools consisted of
                              I.M.A.G.E. clones 297480-302087, 682632-687239,
                              726408-728711, and 729096-731399. Subtraction by Bento
                              Soares and M. Fatima Bonaldo."
BASE COUNT       100 a       106 c       115 g       109 t
ORIGIN

Alignment Scores:
Pred. No.:               1.75e-14      Length:            430
Score:                   218.50        Matches:           47
Percent Similarity:      64.89%        Conservative:      14
Best Local Similarity:   50.00%        Mismatches:        32
Query Match:             42.51%        Indels:            1
DB:                      9             Gaps:              1

US-10-045-180A-3 (1-94) x AI684407 (1-430)

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
         ||||||||||||||||   :::|||||  ||||||||||||||||||||||||||||
Db  404 ATGAGGACCCTCGCCATCCTTGCCATCTTCCCTGCGTGGGCCCTGCAGGCCTGAG 345

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp  39
         ||||||||||||  |||:::   ::|||     ||:::|||||||||||||||||||||
Db  344 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGGAGCAGATTGCAGCGGACATC 285

Qy   40 GlnAspValValIleTyrPheSerGlyAspSerCysSerLeuGlnValProGlySer  59
         ||||||||||||:::  |||  |||  :::  |||||||||||||||||||||||||
Db  284 CCAGGAGTGGTTGTTTCCCTGCGGACGAGCTTGGCTCCAAAGCATCCAGGCTCA 225

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGluGlyHisLeuGly  79
         ||| :::||||||||||:::    ::: |||||||||  |||||||
Db  224 AGGAAAAACATGCCCTGCTATTACCAGCGTGCATTCTGCTGC 165

Qy   80 GlyThrCysPheIleLeuGlyLeuGluArgTyrProIleCysCys  93
         |||||   |||||||   :::   ||||||||||||||||
Db  164 GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC 123

RESULT 15
AI806882/c
LOCUS     AI806882            430 bp    mRNA      linear    EST 19-DEC-1999
DEFINITION  wf36d08.x1 Soares NFL T GBC S1 Homo sapiens cDNA clone
            IMAGE:2357679 3' similar to gb:M26602 NEUTROPHIL DEFENSINS 1, 2 AND
            3 PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   AI806882
VERSION     AI806882.1      GI:5393448
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 430)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 286   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1..430
```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2357679"
/lab_host="DH10B"
/clone_lib="Soares_NFL_T_GBC_S1"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHi19W, testis NHT, and B-cell
NCI_CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "

BASE COUNT   99 a   105 c   117 g   109 t
ORIGIN

Alignment Scores:
Pred. No.:   Length: 430
Score: 218.50   Matches: 47
Percent Similarity: 64.89%   Conservative: 14
Best Local Similarity: 50.00%   Mismatches: 32
Query Match: 42.51%   Indels: 1
DB: 9   Gaps: 1

US-10-045-180A-3 (1-94) x AI806882 (1-430)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaAlaGlu 20
            |||||||||||||||| ::::||| || |||||||||||| |||||||||
Db   404 ATGAGGACCCTCGCCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCTGAG 345

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
            |||||||||||||||||||||| |||::: || :::||| ||| |||||
Db   344 CCACTCCAGGCAAGAGCTTGATGAGGTTGCTGCAGCCCCGGAGCAGGATTGCAGCGGACATC 285

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
            ::||||||||| |||||||| |||||| || ||||| || :::||||||||||||
Db   284 CCAGAAGTGGTTGGTTGTTCCCTTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 225

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
            ||| ::: ||||||||||||||| || |||||||| ||||| |||||
Db   224 AGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTAT 165

Qy   80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
            |||||||||| ||| ::: |||
Db   164 GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC 123

Search completed: December 17, 2003, 18:25:42
Job time : 3291.5 secs

```
           GenCore version 5.1.6
       Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:     December 17, 2003, 14:50:33 ; Search time 352.5 Seconds
                                        (without alignments)
                                        719.850 Million cell updates/sec

Title:      US-10-045-180A-3
Perfect score: 514
Sequence:   1 MRTLTLLSAFLLVALQAWAE.........GEHLGGTCFILGGRYPICCY 94

Scoring table: BLOSUM62
            Xgapop  10.0 , Xgapext   0.5
            Ygapop  10.0 , Ygapext   0.5
            Fgapop   6.0 , Fgapext   7.0
            Delop    6.0 , Delext    7.0

Searched:   252756 seqs, 134919017 residues

Total number of hits satisfying chosen parameters:       5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145023_12520/app_query.fasta_1.860
-DB=N Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10045180 @CGN_1_1_835 @runat_17122003_145023_12520 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=30 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -YGAPEXT=7 -DELOP=6 -DELEXT=7

Database :  N Geneseq 19Jun03:*
    1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
    2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
    3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
    4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
    5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
    6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
    7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
    8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
    9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
   10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
   11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
   12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
   13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
   14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
   15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
   16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
   17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
   18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
   19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
   20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
   21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
   22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
   23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
   24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
   25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*
```

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	514	100.0	453	20	AAX26697	cDNA sequence enco
2	514	100.0	453	20	AAX26698	cDNA sequence enco
3	393.5	76.6	4415	20	AAX26696	Nucleotide sequenc
4	227	44.2	860	25	ABX63473	Human cDNA #473 di
5	220	42.8	452	14	AAQ53217	Sequence encoding
6	220	42.8	452	25	ACA56600	Human signalling p
7	218.5	42.5	478	21	AAC04468	Human secreted pro
8	218.5	42.5	498	12	AAQ14908	Chronic myelogenou
9	218.5	42.5	498	24	ABQ80601	Human defensin 1 c
10	218.5	42.5	514	24	ABN59722	Novel human coding
11	218.5	42.5	652	24	ABZ11202	Human polynucleoti
12	218.5	42.5	664	24	ABZ11183	Human polynucleoti
13	216.5	42.1	464	21	AAF20911	Human defensin 3 p
14	216.5	42.1	464	21	AAA34789	Human adenosine re
15	216.5	42.1	644	21	AAF21446	Human defensin 3 p
16	216.5	42.1	209273	21	AAF21437	Human factor-relat
17	209.5	40.8	299	16	AAT26106	Human gene signatu
18	207	40.3	424	14	AAQ53216	Sequence encoding
19	207	40.3	449	24	ABQ60783	Human defensin 5 D
20	200	38.9	411	24	ABQ60009	Human colon cancer
21 c	198	38.5	586	24	ABQ60212	Human colon cancer
22	197.5	38.4	543	25	ABZ36767	Human GENSET codin
23	196	38.1	514	22	AAH57427	Human intestine ce
24	175.5	34.1	496	25	AAD51767	Human retrocyclin
25	173.5	33.8	420	17	AAT30745	Mouse cryptdin 7 c
26	171.5	33.4	401	17	AAT30753	Mouse cryptdin 15
27	168.5	32.8	388	17	AAT30751	Mouse cryptdin 13
28	166.5	32.4	377	17	AAT30747	Mouse cryptdin 9 c
29	165.5	32.2	422	17	AAT30740	Mouse cryptdin 2 c
30	165.5	32.2	500	22	AAC66924	Rhesus macaque the
31	165.5	32.2	500	25	AAD51768	Macaca mulatta RTD
32	164.5	32.0	422	17	AAT30741	Mouse cryptdin 3 c
33	163.5	31.8	294	22	AAH57220	Human intestine sp
34	163.5	31.8	422	17	AAT30739	Mouse cryptdin 1 c
35	160	31.1	365	17	AAT30742	Mouse cryptdin 4 c
36 c	160	31.1	412	24	ABK39009	cDNA encoding lung
37 c	160	31.1	412	25	ACA11338	Human lung adenoca
38 c	160	31.1	412	25	ACA02524	Lung cancer therap
39	159.5	31.0	391	17	AAT30754	Mouse cryptdin 16
40	158.5	30.8	375	17	AAT30748	Mouse cryptdin 10
41	157.5	30.6	419	17	AAT30729	Mouse cryptdin 3 g
42	157.5	30.6	422	17	AAT30744	Mouse cryptdin 6 c
43	157.5	30.6	495	22	AAC66925	Rhesus macaque the
44	157.5	30.6	495	25	AAD51769	Macaca mulatta RTD
45	154.5	30.1	421	17	AAT30743	Mouse cryptdin 5 c

ALIGNMENTS

```
RESULT 1
AAX26697
ID   AAX26697 standard; cDNA; 453 BP.
XX
AC   AAX26697;
XX
DT   18-JUN-1999 (first entry)
XX
DE   cDNA sequence encoding human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis; ss.
XX
OS   Homo sapiens.
XX
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

Left column:

PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;   97FR-0010823.
XX
PR   29-AUG-1997;   97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
DR   P-PSDB; AAY01604.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Disclosure; Fig 3; 56pp; French.
XX
CC   The present sequence encodes human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 453 BP; 107 A; 121 C; 97 G; 128 T; 0 other;

Alignment Scores:
Pred. No.:   7.1e-54   Length: 453
Score: 514.00   Matches: 94
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 20   Gaps: 0

US-10-045-180A-3 (1-94) x AAX26697 (1-453)

Qy   1 MetArgThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaAlaTrpAlaGlu 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   52 ATGAGGACCCTCACCCTCCTCTCGGCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCAGAG 111

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnLysProProAlaAspGln 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   112 CCGCTTCCAGGCAAGAGCTCATGAGATGCCCAGCCCAGAAGCAGCCTCCAGCAGATGACCAG 171

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   172 GATGTGGTCATTTACTTTCAGGAGAGATGACAGCTGCTCTTCTTCAGGTTCCAGGCTCAACA 231

Qy   61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   232 AAGGGCTTGATCTGCCATTGCAGAGTACTATACTGCATTTTTGGAGAAACATCTTCGTGGG 291

Qy   81 ThrCysPheLeuGluGlyArgTyrProIleCysCysTyr 94
         ||||||||||||||||||||||||||||||||||||
Db   292 ACCTGCTTCATCCTTGGTGAACGTCACCCAATCTGCTGCTAC 333

RESULT 2

Right column:

ID   AAX26698 standard; cDNA; 453 BP.
XX
AC   AAX26698;
XX
DT   18-JUN-1999 (first entry)
XX
DE   cDNA sequence encoding human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;   97FR-0010823.
XX
PR   29-AUG-1997;   97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Claim 5; Fig 4; 56pp; French.
XX
CC   The present sequence encodes human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 453 BP; 107 A; 121 C; 97 G; 128 T; 0 other;

Alignment Scores:
Pred. No.:   7.1e-54   Length: 453
Score: 514.00   Matches: 94
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 20   Gaps: 0

US-10-045-180A-3 (1-94) x AAX26698 (1-453)

Qy   1 MetArgThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaAlaTrpAlaGlu 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   52 ATGAGGACCCTCACCCTCCTCTCGGCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCAGAG 111

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnLysProProAlaAspGln 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Pred. No.:              8.7e-38      Length:        4415
Score:                  393.50      Matches:         94
Percent Similarity:     32.41%      Conservative:     0
Best Local Similarity:  32.41%      Mismatches:       0
Query Match:            76.56%      Indels:         196
                        20          Gaps:             1
DB:
```

US-10-045-180a-3 (1-94) x AAX26696 (1-4415)

```
Qy      1  MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaAlaGlnAlaTrpAlaGlu  20
Db   3406  ATGAGGACCCTCACCCTCCTCCTGCCTTTCTCCCGGTGGCCCTTCAGGCCTGGGCAGAG  3465

Qy     21  ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln  40
Db   3466  CCGTTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGACCAG  3525

Qy     41  AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGln-------------  55
Db   3526  GATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGTGAGAGA  3585

Qy     55  -------------------------------------------------------  55
Db   3586  TGCCAGCATGCAGAGCTACAGACTAGACAGGAAGGACAGGAGCTCTGGAATTGGAT  3645

Qy     55  -------------------------------------------------------  55
Db   3646  CTCAGTGGCAGATGTCACTTAGGTGGCTATACTTAACATCTCTGGTCCTGGATTTTCTCA  3705

Qy     55  -------------------------------------------------------  55
Db   3706  TATCTAAATGAATAGAGAACCAAAGAAATCTAAGAGATTTTCTTCTCCAAAAACTTG  3765

Qy     55  -------------------------------------------------------  55
Db   3766  ATTCCAAGATATGACTGTGAAATTCACTAGATTTAAGATATAAGGAGATGCTACCTAGTT  3825

Qy     55  -------------------------------------------------------  55
Db   3826  CCTTCTGGAGCCAGCAGACAAACAAGCTTAAGTATATAGAGAAAATATTTCACCCTGTCTATAT  3885

Qy     55  -------------------------------------------------------  55
Db   3886  AGGAGGTTTTAGAACCTGGAGAGGAGCCTAAGATGTGTTCAGGTGTGTGTGATGGGC  3945

Qy     55  -------------------------------------------------------  55
Db   3946  AGGAATGCAGAAAGTGAAGCAAAGGAAGAATGAGTCTCGAATCCTGTGTGACCAGCACTG  4005

Qy     55  -------------------------------------------------------  55
Db   4006  CTCTGTGTATTATTCCTATTGACTGAGATTGTTGTGTACCGGCTGTAATACAGCCAA  4065

Qy     55  -------------------------------------------------------  55
Db   4066  CATCACTCATCAGCCAACATGTGACTTCTCCAAGATTCCCTTACCACCCACTGCTGNAC  4125

Qy     56  -----------------------------------Val-ProGlySerThrLysGlyLeuIleCy  65
Db   4126  CCCGTACTCAGTTTCTGATGCTCTTCTGGTCCCCAGGCTCAACAAGGCTTGATCTG  4185

Qy     65  sHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyTyrCysPheIleLe  85
Db   4186  CCATTGCAGAGTACTATACTGCATTTTTGGAGAACATCTTGGTGGGACCTGCTTCATCCT  4245

Qy     85  uGlyGluArgTyrProIleCysCysTyr  94
Db   4246  TGGTGAACGCTACCCAATCTGCTGCTAC  4273
```

```
Db    112  CCGGTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGACCAG  171
Qy     41  AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr  60
Db    172  GATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGCTCAACA  231

Qy     61  LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly  80
Db    232  AAGGGCTTGATCTGCCATTGCAGAGTACTATACTGCATTTTTGGAGAACATCTTGGTGGG  291

Qy     81  ThrCysPheIleLeuGlyGluArgTyrProIleCysCysTyr  94
Db    292  ACCTGCTTCATCCTTGGTGAACGCTACCCAATCTGCTGCTAC  333
```

RESULT 3
AAX26696
ID   AAX26696 standard; DNA; 4415 BP.
XX
AC   AAX26696;
XX
DT   18-JUN-1999 (first entry)
XX
DE   Nucleotide sequence of human defensin (Def-X).
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997; 97FR-0010823.
XX
PR   29-AUG-1997; 97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Claim 5; Fig 2; 56pp; French.
XX
CC   The present sequence encodes human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 4415 BP; 1128 A; 1109 C; 912 G; 1252 T; 14 other;

Alignment Scores:

RESULT 4
ABX63473
ID   ABX63473 standard; cDNA; 860 BP.
XX

Thu Dec 18 06:09:39 2003

us-10-045-180a-3.rng

```
AC   ABX63473;
XX
DT   26-FEB-2003  (first entry)
XX
DE   Human cDNA #473 differentially expressed in activated vascular tissue.
XX
KW   Human; gene; ss; vascular tissue; cytostatic; atherosclerosis;
KW   cardiant; hypotensive; antidiabetic; gynaecological; vasotropic;
KW   cerebroprotective; gene therapy; vascular disease; cancer; coronary;
KW   artery disease; hypertension; diabetes; pre-eclampsia; restenosis;
KW   ischaemia-reperfusion injury; stroke;
XX
OS   Homo sapiens.
XX
PN   US2002137081-A1.
XX
PD   26-SEP-2002.
XX
PF   08-JAN-2002; 2002US-0044090.
XX
PR   28-JUL-2000; 2000US-222469P.
PR   08-JAN-2001; 2001US-260483P.
XX
PA   (BAND/) BANDMAN O.
XX
PI   Bandman O;
XX
DR   WPI; 2003-110597/10.
XX
PT   Combination for diagnosing, staging, treating, or monitoring the
PT   progression of treatment of a vascular disease, e.g. atherosclerosis,
PT   comprises several cDNAs that are differentially expressed in activated
PT   vascular tissue -
XX
PS   Claim 1; Page -; 18pp; English.
XX
CC   This invention relates to a combination comprising several cDNAs that
CC   are differentially expressed in activated vascular tissue. The invention
CC   also discloses a high throughput method for detecting differentially
CC   expressed cDNAs in a sample. The cDNAs of the invention may have
CC   antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC   gynaecological; vasotropic and cerebroprotective activities and may be
CC   used in gene therapy. The cDNAs of the invention may be used in a
CC   high-throughput methods for detecting differential expression of one or
CC   more cDNAs in a sample, or screening several molecules or compounds to
CC   identify a molecule or compound that specifically binds a cDNA of the
CC   invention. A protein encoded by the cDNA may be used to screen several
CC   molecules or compounds to identify a ligand that specifically binds to
CC   the protein, or to produce or purify an antibody to the protein that can
CC   be used to detect a protein in a sample or purify a natural or
CC   recombinant protein from a sample. The nucleotides may be useful for
CC   diagnosing, staging, treating, or monitoring the progression of
CC   treatment of a vascular disease, e.g. atherosclerosis, cancer, coronary
CC   artery disease, hypertension, diabetes, pre-eclampsia, ischaemia-
CC   reperfusion injury, restenosis, or stroke. The cDNAs can also be used
CC   for large-scale genetic or gene expression analysis of several new
CC   nucleic acid molecules. Antibodies to the proteins encoded by the
CC   cDNAs are useful for diagnosing pre-pathologic disorders, and chronic
CC   or acute diseases associated with abnormalities in the expression,
CC   amount or distribution of the protein. The present sequence
CC   represents a cDNA of the invention that is differentially expressed in
CC   activated vascular tissue.
CC   Note: The sequence data for this patent did not form part of the
CC   specification, but was obtained in electronic format directly from USPTO
CC   at http://seqdata.uspto.gov/sequence.html?DocID=2002013708l.
XX
SQ   Sequence 860 BP; 216 A; 211 C; 192 G; 241 T; 0 other;

Alignment Scores:
Pred. No.:                          Length:         860
Score:             227.00          Matches:         46
Percent Similarity: 64.52%         Conservative:    14
Best Local Similarity: 49.46%      Mismatches:      33
```

```
Query Match:        44.16%       Indels:    0
DB:                 25           Gaps:      0

US-10-045-180A-3 (1-94) x ABX63473 (1-860)

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaAlaGlu 20
          |||       ::  ||||||||  |||||||||||||
Db   76 ATGAGGATTATCGCCCTCCTCGTGTATTCCTTGGTAGCGCTCCAGGTCCGGCAGGC 135

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
          ||||||||||||| |||::|| |||::|||  ::|||
Db  136 CCACTCCAGGCAAGAGGTGATGAGGCTCCAGGCAGGCAGCGGTGGGCCAGAAGACCAG 195

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
          |||::: ||||  :||::| :::||
Db  196 GACATATCTATTTCCTTGCATGGGATAAAAGCTCTGCTTCCTTCAGGTTCAGGCTCAACA 255

Qy   61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly 80
          :::|||:::::|||||||:::::|||  |||
Db  256 AGGGGCATGGTCTGCTTGCAGATTAGTATTCTGCCGGCAACAGAACTTCGTGTTGGG 315

Qy   81 ThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
          |||  :::  |||
Db  316 AACTGCCTCATTGGTGGTGTGGAGTTTCACATACTGCTGC 354
```

```
RESULT 5
AAQ53217
ID   AAQ53217 standard; cDNA; 452 BP.
XX
AC   AAQ53217;
XX
DT   25-MAR-2003  (updated)
DT   17-JUN-1994  (first entry)
XX
DE   Sequence encoding gastrointestinal defensin (GID) peptide called
DE   human defensin 6.
XX
KW   Gastrointestinal defensin peptide; GID; pharmaceutical; Paneth
KW   cell; antimicrobial; anti-inflammatory; diagnosis;
KW   contact disinfectant; ss.
XX
OS   Homo sapiens.
XX
FH   Key        Location/Qualifiers
FT   CDS        19..321
FT              /*tag= a
XX
PN   WO9324513-A1.
XX
PD   09-DEC-1993.
XX
PF   18-MAY-1993; 93WO-US04740.
XX
PR   22-MAY-1992; 92US-0888232.
XX
PA   (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.
XX
PI   Bevins CL, Jones DE;
XX
DR   WPI; 1993-405719/50.
DR   P-PSDB; AAR44819.
XX
PT   Gastrointestinal defensin peptide(s) - useful as antimicrobial
PT   and anti-inflammatory agents and for detecting gastrointestinal
PT   disorders
XX
PS   Claim 4; page 48-49; 97pp; English.
XX
CC   A probe (D5' oligo) having the sequence given in AAQ53219 (claimed)
CC   was used to screen a population of clones to identify novel defensin
CC   defensin peptides. Pref. the libraries were human genomic and cDNA
CC   libraries. Hybridisation and partial sequence analysis of the
CC   identified clones contained previously characterised myeloid derived
```

Left column

```
CC   defensin sequences as well as new defensin related sequences. Two
CC   clones expressing new defensin related sequences were extensively
CC   characterised and found to contain genes selectively expressed in
CC   Paneth cells of the small intestin. These Paneth cell-derived
CC   defensins are designated human defensin 5 and human defensin 6 and
CC   are referred to as gastrointestinal defensin peptides. Comparison
CC   of the deduced AA sequences of the defensin 5 cDNA with the
CC   previously reported preprodefensin shows significant similarity.
CC   The deduced AA sequence of defensin 6 cDNA has features similar to
CC   defensin 5 and the previously reported preprodefensins.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 452 BP; 123 A; 118 C; 98 G; 113 T; 0 other;

Alignment Scores:
Pred. No.:
Score:                 8.96e-18          Length:         452
Percent Similarity:    220.00            Matches:        49
Best Local Similarity: 63.00%            Conservative:   14
Query Match:           49.00%            Mismatches:     29
DB:                    42.80%            Indels:         8
                       14                Gaps:           2

US-10-045-180A-3 (1-94) x AAQ53217 (1-452)

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu   20
        |||||||||||||||||||||||||||:::||||||
Db   19 ATGAGAACCCTCACCATCCTCACTGCTGTTCTCCTTGTGGCCCTCCAGGCCAAGGCTGAG   78

Qy   21 ----------------ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLys         33
                        ||||||||||:||||||||||||||||
Db   79 CCACTCCAAGCTGAGGATGATCCACTGCAGGCAAAAGCTTATGAGGCTGCCCAGGAG       138

Qy   34 GlnProProAlaAspAspGlnAsnValValIleTyrPheSerGlyAspAspSerCysSer    53
        |||       :::    |||||:||        ||
Db  139 CAGCCTGGGCAAATGACCAGGACCTTGCCGTCTCCTTTGCAGAGGATGCAAGCTCAAGT    198

Qy   54 LeuGlnValProGlySerThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIle    73
        ||:::  ||||||||||||||||
Db  199 CTTAGAGCTTTGGGGCTCAACAAGGCTTTCACTTGCCATTGCAGA--AGGTCCTGTTAT   255

Qy   74 PheGlyGluHisLeuGlyGlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys    93
        |||:::          :::::||
Db  256 TCAACAGAATATTCCTATGGGACCTGCACTGTCATGGGTATTAACCACAGATTCTGCTGC   315

RESULT 6
ACA56600
ID   ACA56600 standard; cDNA; 452 BP.
XX
AC   ACA56600;
XX
DT   06-JUN-2003  (first entry)
XX
DE   Human signalling pathway polynucleotide probe SEQ ID NO 1198.
XX
KW   Human; probe; ss; array element; Parkinson's disease;
KW   signalling pathway population; cancer; adenocarcinoma; leukaemia;
KW   immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
XX
OS   Homo sapiens.
XX
PN   US6500938-B1.
XX
PD   31-DEC-2002.
XX
PF   30-JAN-1998;  98US-0016434.
XX
PR   30-JAN-1998;  98US-0016434.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Au-Young J,  Seilhamer JJ;
XX
DR   WPI; 2003-352189/33.
```

Right column

```
XX
PT   Combination of polynucleotide probes, useful as array elements in a
PT   microarray for monitoring the expression of a number of target
PT   polynucleotides -
XX
PS   Claim 1; SEQ ID NO 1198; 65pp; English.
XX
CC   The invention relates to a combination which, comprises a number of
CC   polynucleotide probes comprising a sequence selected from one of the 1490
CC   sequences mentioned in the specification. The combination is useful as an
CC   array element in a microarray for monitoring the expression of a number
CC   of target polynucleotides. The microarray is particularly useful in the
CC   diagnosis and treatment of cancer and immunopathology and neuropathology.
CC   The microarray is useful in diagnostics and treatment regimens, drug
CC   discovery and development, toxicological and carcinogenicity studies,
CC   forensics and pharmacogenomics. The microarray is also useful for
CC   monitoring progression of diseases and for developing sophisticated
CC   profiles for the effects of currently available therapeutic drugs. The
CC   combination is also useful for purifying a subpopulation of mRNAs, cDNAs
CC   and genomic fragments and in research and diagnostic applications. The
CC   array can detect changes in expression in a large number of genes coding
CC   for different signaling pathway populations which can be used to diagnose
CC   various diseases including cancer e.g. adenocarcinoma and leukaemia,
CC   immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
CC   and Parkinson's disease. The present sequence represents a polynucleotide
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=06500938B1.
XX
SQ   Sequence 452 BP; 124 A; 118 C; 98 G; 112 T; 0 other;

Alignment Scores:
Pred. No.:                               Length:         452
Score:                 8.96e-18          Matches:        49
Percent Similarity:    220.00            Conservative:   14
Best Local Similarity: 63.00%            Mismatches:     29
Query Match:           49.00%            Indels:         8
DB:                    42.80%            Gaps:           2
                       25

US-10-045-180A-3 (1-94) x ACA56600 (1-452)

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaLeuAlaAlaGlu   20
        |||||||||||||||||||||||||||::::||||||
Db   19 ATGAGAACCCTCACCATCCTCACTGCTGTTCTCCTTGTGGCCCTCCAGGCCAAGGCTGAG   78

Qy   21 ----------------ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLys         33
                        ||||||||||:||||||||||||||||
Db   79 CCACTCCAAGCTGAGGATGATCCACTGCAGGCAAAAGCTTATGAGGCTCCCCAGGAG      138

Qy   34 GlnProProAlaAspAspGlnAsnValValIleTyrPheSerGlyAspAspSerCysSer    53
        |||       :::    |||||:||        ||
Db  139 CAGCCTGGGCAAATGACCAGGACCTTGCCGTCTCCTTTGCAGAGGATGCAAGCTCAAGT    198

Qy   54 LeuGlnValProGlySerThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIle    73
        ||:::  ||||||||||||||||
Db  199 CTTAGAGCTTTGGGGCTCAACAAGGCTTTCACTTGCCATTGCAGA--AGGTCCTGTTAT   255

Qy   74 PheGlyGluHisLeuGlyGlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys    93
        |||:::          ::::::||
Db  256 TCAACAGAATATTCCTATGGGACCTGCACTGTCATGGGTATTAACCACAGATTCTGCTGC   315

RESULT 7
AAC04468
ID   AAC04468 standard; cDNA; 478 BP.
XX
AC   AAC04468;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 8543.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
```

KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS Claim 1; SEQ ID 8543; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 478 BP; 117 A; 129 C; 127 G; 101 T; 4 other;

Alignment Scores:
Pred. No.:
Score: 1.47e-17
Percent Similarity: 218.50
Best Local Similarity: 64.89%
Query Match: 50.00%
DB: 42.51%
          21

Length: 478
Matches: 47
Conservative: 14
Mismatches: 32
Indels: 1
Gaps: 1

US-10-045-180A-3 (1-94) x AAC04468 (1-478)

Qy 1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
    |||||||||||||| ::||||:||||||||||||||||||||||||||
Db 151 ATGAGGACCCTCGCCATCCTTGCTGCCATTCTCCTGTGGCCCTGCAGGCCCAGGCTGAG 210

Qy 21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysProProAlaAspAsp 39
    |||||||| ::|||||||:|||| :::|| |||||
Db 211 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGAGCAGATTGCAGCGGACATC 270

Qy 40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
    ::||||||||| ::: |||:|||||| ::: ||||||||
Db 271 CCAGAAGTGGTTGTTTCCCTTGCATGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 330

Qy 60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
    ||| ||||||||||||:||||||||||||||:||||||||||
Db 331 AGGAAAAACATGGMCTGCTATTGCAGAATACCAGCCTGCATTGCAGGAGAACGTCGCTAT 390

Qy 80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
    |||| :::
Db 391 GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC 432

ID AAQ14908 standard; cDNA; 498 BP.
XX
AC AAQ14908;
XX
DT 25-FEB-1992 (first entry)
XX
DE Chronic myelogenous leukaemia-derived mrs cDNA.
XX
KW CML; Philadelphia chromosome; myeloid-related sequence;
KW chromosome 8; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   16..378
FT   /*tag= a
FT   /note= "see comments"
FT polyA_signal 479..484
FT   /*tag= b
XX
PN US5066792-A.
XX
PD 19-NOV-1991.
XX
PF 02-DEC-1988; 88US-0279451.
XX
PR 02-DEC-1988; 88US-0279451.
PR 28-SEP-1984; 84US-0655942.
PR 04-SEP-1987; 87US-0094099.
XX
PA (TEXA) UNIV OF TEXAS.
XX
PI Saunders GF, Mars WM;
XX
DR WPI; 1991-361533/49.
DR P-PSDB; AAR15222.
XX
PT Gene probe for chronic myelogenous leukaemia - esp. useful for
PT distinguishing myeloblastic crisis from lymphoblastic crisis
PS Disclosure; Fig 2; 13pp; English.
XX
CC The precise position of the CDS is unsure due to the 13 unidentified
CC nucleotides at the 5' end of the sequence. The most likely position
CC for initiation of translation occurs at nucleotide 94 and the open
CC reading frame following this point is claimed. The first 18 codons
CC of the claimed ORF (excluding the ATG) encode a sequence of amino
CC acids that resembles other leader sequences and their cleavage sites.
CC The mrs sequence was isolated as clone C-A3 (ATCC 39868) from a cDNA
CC library constructed from total RNA from a patient suffering from
CC chronic phase, Ph'-positive CML. The mrs is localised to human
CC chromosome 8, specifically to bands 8q21.1-23.
XX
SQ Sequence 498 BP; 118 A; 134 C; 122 G; 111 T; 13 other;

Alignment Scores:
Pred. No.:
Score: 1.55e-17
Percent Similarity: 218.50
Best Local Similarity: 64.89%
Query Match: 50.00%
DB: 42.51%
          12

Length: 498
Matches: 47
Conservative: 14
Mismatches: 32
Indels: 1
Gaps: 1

US-10-045-180A-3 (1-94) x AAQ14908 (1-498)

Qy 1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
    |||||||||||||| ::||||:||||||||||||||||||||||||||
Db 94 ATGAGGACCCTCGCCATCCTTGCTGCCATTCTCCTGTGGCCCTGCAGGCCCAGGCTGAG 153

Qy 21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysProProAlaAspAsp 39
    |||||||| ::|||||||:|||| :::|| |||||
Db 154 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGAGCAGATTGCAGCGGACATC 213

RESULT 8
AAQ14908

```
Qy    40  GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
              ::|||||||||  :::  ||:  :::  ||||||
Db   214  CCAGAAGTGGTTGTTTCCCTTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 273

Qy    60  ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyHisLeuGly 79
              |||  :::  ||||||:::
Db   274  AGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGGAACGTCGCTAT 333

Qy    80  GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
              :::  |||||
Db   334  GGAACCTGCATCTACCAGGGAGACTCTGGGCATTCTGCTGC 375

RESULT 9
ABQ80601
ID   ABQ80601 standard; DNA, 498 BP.
XX
AC   ABQ80601;
XX
DT   11-NOV-2002  (first entry)
XX
DE   Human defensin 1 coding sequence.
XX
KW   Antiallergic; allergic disease; allergy; defensin 1; atopic dermatitis;
KW   human; gene; ds.
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            94..378
FT                  /*tag= a
FT                  /product= "Human defensin 1"
XX
PN   WO20064832-A1.
XX
PD   22-AUG-2002.
XX
PF   13-FEB-2002; 2002WO-JP01193.
XX
PR   14-FEB-2001; 2001JP-0036362.
XX
PA   (GENO-) GENOX RES INC.
PA   (NIGE-) JAPAN GEN AGENCY NATION.
XX
PI   Sugita Y, Heishi M, Kagaya S, Gunji S, Saito H;
XX
DR   WPI; 2002-674884/72.
DR   P-PSDB; ABB98494.
XX
PT   Examining allergic diseases by changes in expression levels of
PT   defensin1 gene in peripheral blood monocytes as indication, also
PT   applicable in screening compounds for treating of allergic diseases
PT   e.g. atopic dermatitis -
XX
PS   Disclosure; Page 45-46; 57pp; Japanese.
XX
CC   The present invention relates to a method for examining allergic
CC   diseases. The method comprises: (a) determining the expression level of
CC   defensin 1 gene in the biological sample from a patient; and (b)
CC   comparing the expression level with that in the sample of a healthy
CC   individual. The method is for examining allergic diseases particularly
CC   atopic dermatitis and its diagnosis, which is also applicable in
CC   screening candidate compounds for remedies. The present sequence is the
CC   coding sequence for human defensin 1.
XX
SQ   Sequence 498 BP; 118 A; 134 C; 122 G; 111 T; 13 other;

Alignment Scores:
Pred. No.:          1.55e-17        Length:        498
Score:              218.50          Matches:        47
Percent Similarity: 64.89%          Conservative:   14
Best Local Similarity: 50.00%       Mismatches:     32
Query Match:        42.51%          Indels:          1
DB:                 24              Gaps:            1
```

```
US-10-045-180A-3 (1-94) x ABQ80601 (1-498)

Qy     1  MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
              |||||||||  ::  |||  |||  ||||||||||||||||||||||
Db    94  ATGAGGACCCTGCACATCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCTGAG 153

Qy    21  ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
              |||||||||||  ||||||  :::  |||||
Db   154  CCACTTCAGGCAAGAGCTCATGAGGTTGCTGCAGCCCCGGAGCAGATTGCAGCCGGACATC 213

Qy    40  GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
              ::|||||||||  :::  ||:  :::  ||||||
Db   214  CCAGAAGTGGTTGTTTCCCTTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 273

Qy    60  ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyHisLeuGly 79
              |||  :::  ||||||:::
Db   274  AGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGGAACGTCGCTAT 333

Qy    80  GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
              :::  |||||
Db   334  GGAACCTGCATCTACCAGGGAGACTCTGGGCATTCTGCTGC 375

RESULT 10
ABN59722
ID   ABN59722 standard; cDNA; 514 BP.
XX
AC   ABN59722;
XX
DT   28-JUN-2002  (first entry)
XX
DE   Novel human coding sequence SEQ ID NO: 133.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag; gene; ss.
OS   Homo sapiens.
XX
PN   WO20022660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US26015.
PR   11-SEP-2000; 2000US-0659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   P-PSDB; ABB97309.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Claim 1; SEQ ID NO 133; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC   e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a coding sequence of the
CC   invention.
XX
```

platelet or coagulation disorders -

PT
XX
PS　Claim 1; SEQ ID NO 84; 1012pp + Sequence Listing; English.
XX
CC　The invention relates to an isolated polynucleotide (I) comprising a
CC　nucleotide sequence selected from any of 948 sequences
CC　(ABZ11119-ABZ12066) or their mature protein coding portion, active domain
CC　coding protein or complementary sequences. The polynucleotides are useful
CC　for identifying expressed genes or for physical mapping of human genome.
CC　The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC　weight markers, as a food supplement, for generating antibodies, in
CC　medical imaging, screening and diagnostic assays and for treating
CC　cell-proliferative disorders (cancer), neurodegenerative diseases
CC　(Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC　sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC　disorders, platelet or coagulation disorders, wound, burns, incision,
CC　ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC　parasitic), arthritis, etc.
CC　Note: The sequence data for this patent did not form part of the printed
CC　specification, but was obtained in electronic format directly from WIPO
CC　at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ　Sequence 652 BP; 146 A; 185 C; 177 G; 144 T; 0 other;

Alignment Scores:
Pred. No.:　　　　　　　　　　　　Length:　　　　652
Score:　　　　　　　2.21e-17　　　　Matches:　　　　47
Percent Similarity:　64.89%　　　　Conservative:　14
Best Local Similarity:　50.00%　　　　Mismatches:　　32
Query Match:　　　　42.51%　　　　Indels:　　　　1
DB:　　　　　　　　　　24　　　　　Gaps:　　　　　1

US-10-045-180A-3 (1-94) x ABZ11202 (1-652)

Qy　　1　MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu　20
　　　　　||||||||||||　　　:::||||||||||||||||　||||||
Db　269　ATGAGGACCCTGCCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCCCAGGCTGAG　328

Qy　21　ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp　39
　　　　　||||||||||||||　　|||　　:::||　　　|||||||||||
Db　329　CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCCGGAGCAGCAGATC　388

Qy　40　GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer　59
　　　　　:::||||||||||||　　||||||　　:::　:::|||||||||||
Db　389　CCAGAAGTGGTTGTTTCCCTTGCATGGGACGGAAAGCTTGGCTCCAAAGCATCCAGGCTCA　448

Qy　60　ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly　79
　　　　　||||||||||||||||||||||　　:::　||||||||||||
Db　449　AGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTAT　508

Qy　80　GlyThrCysPheLeuGlyGluArgTyrProIleCysCys　93
　　　　　||||||||||　　　　:::　|||
Db　509　GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC　550

RESULT 12
ABZ11183
ID　ABZ11183 standard; cDNA; 664 BP.
XX
AC　ABZ11183;
XX
DT　20-JAN-2003　(first entry)
XX
DE　Human polynucleotide SEQ ID NO 65.
XX
KW　Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW　cell-proliferative disorder; neurodegenerative disease; bacterial;
KW　Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW　multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW　arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW　antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW　haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW　antiarthritic; gene; ss.
XX

SQ　Sequence 514 BP; 138 A; 142 C; 118 G; 116 T; 0 other;

Alignment Scores:
Pred. No.:　　　　　　　　　　　　Length:　　　　514
Score:　　　　　　　1.62e-17　　　　Matches:　　　　47
Percent Similarity:　64.89%　　　　Conservative:　14
Best Local Similarity:　50.00%　　　　Mismatches:　　32
Query Match:　　　　42.51%　　　　Indels:　　　　1
DB:　　　　　　　　　　24　　　　　Gaps:　　　　　1

US-10-045-180A-3 (1-94) x ABN59722 (1-514)

Qy　　1　MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu　20
　　　　　||||||||||||　　　:::||||||||||||||||　||||||
Db　97　ATGAGGACCCTGCCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCCCAGGCTGAG　156

Qy　21　ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp　39
　　　　　||||||||||||||　　|||　　:::||　　　|||||||||||
Db　157　CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCCGGAGCAGCAGATTGCAGCGGACATC　216

Qy　40　GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer　59
　　　　　:::||||||||||||　　||||||　　:::　:::|||||||||||
Db　217　CCAGAAGTGGTTGTTTCCCTTGCATGGGACGGAAAGCATCCAAAGCATCCAGGCTCA　276

Qy　60　ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly　79
　　　　　||||||||||||||||||||||　　:::　||||||||||||
Db　277　AGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTAT　336

Qy　80　GlyThrCysPheLeuGlyGluArgTyrProIleCysCys　93
　　　　　||||||||||　　　　:::　|||
Db　337　GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC　378

RESULT 11
ABZ11202
ID　ABZ11202 standard; cDNA; 652 BP.
XX
AC　ABZ11202;
XX
DT　20-JAN-2003　(first entry)
XX
DE　Human polynucleotide SEQ ID NO 84.
XX
KW　Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW　cell-proliferative disorder; neurodegenerative disease; bacterial;
KW　Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW　multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW　arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW　antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW　haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW　antiarthritic; gene; ss.
XX
OS　Homo sapiens.
XX
PN　WO200270539-A2.
XX
PD　12-SEP-2002.
XX
PF　05-MAR-2002; 2002WO-US05095.
XX
PR　05-MAR-2001; 2001US-0799451.
XX
PA　(HYSE-) HYSEQ INC.
XX
PI　Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI　Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI　Wehrman T, Wang J, Wang D, Drmanac RT;
XX
DR　WPI; 2002-759812/82.
DR　P-PSDB; ABP68985.
XX
PT　New polynucleotides comprising sequences assembled from expressed
PT　sequence tags (ESTs), useful for treating cell-proliferative,
PT　neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or

OS   Homo sapiens.
XX
PN   WO20020539-A2.
XX
PD   12-SEP-2002.
XX
PF   05-MAR-2002; 2002WO-US05095.
XX
PR   05-MAR-2001; 2001US-0799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren P;
PI   Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI   Wehrman T, Wang J, Wang D, Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   P-PSDB; ABP68966.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT   platelet or coagulation disorders -
XX
PS   Claim 1; SEQ ID NO 65; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences
CC   (ABZ11119-ABZ12066) or their mature protein coding portion, active domain
CC   coding protein or complementary sequences. The polynucleotides are useful
CC   for identifying expressed genes or for physical mapping of human genome.
CC   The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC   weight markers, as a food supplement, for generating antibodies, in
CC   medical imaging, screening and diagnostic assays and for treating
CC   cell-proliferative disorders (cancer), neurodegenerative diseases
CC   (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC   sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC   disorders, platelet or coagulation disorders, wound, burns, incision,
CC   ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC   parasitic), arthritis, etc.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 664 BP; 174 A; 168 C; 161 G; 161 T; 0 other;

Alignment Scores:
Pred. No.:   Length: 664
Score: 2.26e-17 Matches: 47
Percent Similarity: 64.89% Conservative: 14
Best Local Similarity: 50.00% Mismatches: 32
Query Match: 42.51%   Indels: 1
DB: Gaps: 1

US-10-045-180A-3 (1-664) x ABZ11183 (1-664)

Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
          ||||||||||||||||||| :::||| ::||||||||:|:||||||||||
Db   251 ATGAGGACCCTGACCCTGCCATCCTTCCTGCCATTCCTGGTGGCCCTGCAGGCCCAGGCTGAG 310

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLnLysGlnProAlaAspAsp 39
          |||||||||||||||||||||:::|| ||:||||||| :::|||
Db   311 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGAGCAGAGTTGCAGGGACATC 370

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
          :::||||||||:   ::: :|:||||| ::: :::||||||||||
Db   371 CCAGAAGTGGTTGTTTCCCTTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 430

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
          ||| ::: ||:||||||||||| |||||||||||||||||
Db   431 AGGAAAAACATGGCCTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTAT 490

Qy   80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
          ||| ||||||||||||||||

Db   491 GGAACCTGCATCTACCACCAGGAAGAGACTCGGGCATTCTGCTGC 532
          |||||||| ||| ::: ||||||

RESULT 13
AAF20911
ID   AAF20911 standard; DNA; 464 BP.
XX
AC   AAF20911;
XX
DT   14-MAR-2001 (first entry)
XX
DE   Human defensin 3 polynucleotide fragment #2478.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999; 99US-0127958.
XX
PA   (UYBC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 218; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,

CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 464 BP; 110 A; 128 C; 119 G; 107 T; 0 other;

Alignment Scores:
Pred. No.:   2.49e-17   Length: 464
Score: 216.50   Matches: 47
Percent Similarity: 64.89%   Conservative: 14
Best Local Similarity: 50.00%   Mismatches: 32
Query Match: 42.12%   Indels: 1
DB: 21   Gaps: 1

US-10-045-180A-3 (1-94) x AAF20911 (1-464)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
            |||||||||||| :::|||||||||||||||||||||||||| ||||||
Db   90 ATGAGGACCCTGCCATCCTGCCATTCTCCTGGTGGCCCTGCAGGCCCCAGGCTGAG 149

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
            |||||||||| ||| :::|||| |||:::||||||
Db   150 CCACTTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGAGCAGATTGCAGCGGACATC 209

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
            :::|||||| ||| :::|||| ::: :::||||
Db   210 CCAGAAGTGGTTGTTTCCCTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 269

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
            :::|||||| ||| :::||||
Db   270 AGGAAAAACATGGACTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTAT 329

Qy   80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
            ||| :::|||
Db   330 GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC 371

RESULT 14
AAA34789
ID   AAA34789 standard; DNA; 464 BP.
XX
AC   AAA34789;
XX
DT   28-JUL-2000 (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2478.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999; 99WO-US17712.
XX
PR   03-AUG-1998; 98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,

---

PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 612; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 464 BP; 110 A; 128 C; 119 G; 107 T; 0 other;

Alignment Scores:
Pred. No.:   2.49e-17   Length: 464
Score: 216.50   Matches: 47
Percent Similarity: 64.89%   Conservative: 14
Best Local Similarity: 50.00%   Mismatches: 32
Query Match: 42.12%   Indels: 1
DB: 21   Gaps: 1

US-10-045-180A-3 (1-94) x AAA34789 (1-464)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
            |||||||||||| :::|||||||||||||||||||||||||| ||||||
Db   90 ATGAGGACCCTGCCATCCTGCCATTCTCCTGGTGGCCCTGCAGGCCCCAGGCTGAG 149

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
            |||||||||| ||| :::|||| |||:::||||||
Db   150 CCACTTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGAGCAGATTGCAGCGGACATC 209

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
            :::|||||| ||| :::|||| ::: :::||||
Db   210 CCAGAAGTGGTTGTTTCCCTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 269

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
            :::|||||| ||| :::||||
Db   270 AGGAAAAACATGGACTGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTAT 329

Qy   80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
            ||| :::|||
Db   330 GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC 371

RESULT 15
AAF21446
ID   AAF21446 standard; DNA; 644 BP.
XX
AC   AAF21446;
XX
DT   14-MAR-2001 (first entry)
XX
DE   Human defensin 3 polynucleotide fragment #3013.
XX

```
US-10-045-180A-3 (1-94) x AAF21446 (1-644)

Qy    1  MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
         |||||||||||||||||| |||| |||:|||
Db  270  ATGAGGACCCTCGCCATCCTGCTGCCATTCTCCTGGTGSCCCTGCAGGCCCAGGCTGAG  329

Qy   21  ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp  39
         |||||||||||||||||||||         |||  :||:||   ||||||
Db  330  CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCGGAGCAGATTGCAGCGGACATC  389

Qy   40  GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer  59
         ::||||||||::     :: |||:||| ||:||   ::: |||||||
Db  390  CCAGAAGTGGTTGTTTCCCTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA  449

Qy   60  ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyHisHisLeuGly  79
         ::: |||||||::|||||||||::||||
Db  450  AGGAAAAACATGGACTGCTATTGCAGAATACCAGCGTGCATTGCAGAGGAGAACGTCGCTAT  509

Qy   80  GlyThrCysPheIleLeuGlyGlyArgTyrProIleCysCys  93
         |||||||            :::
Db  510  GGAACCTGCATCTACCAGGGAAGACTCTGGGCATTCTGCTGC  551
```

Search completed: December 17, 2003, 15:06:01
Job time : 356.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 15:06:18 ; Search time 325 Seconds
(without alignments)
963.776 Million cell updates/sec

Title: US-10-045-180A-3
Perfect score: 514
Sequence: 1 MRTLTLLSAFLLVALQAWAE.........GEHLGGTCFILGERYPICCY 94

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2211978 seqs, 166101734 residues

Total number of hits satisfying chosen parameters: 4423956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145026_12635/app_query.fasta_1.860
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10045180 @CGN_1_1_174 @runat_17122003_145026_12635
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=30 -DELOP=6 -DELEXT=7
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -XGAPOP=10 -XGAPEXT=0.5

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/pct_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US02_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result   %
No.   Score   Query Match   Length   DB   ID   Description

Flag	No.	Score	% Query Match	Length	DB	ID	Description
	1	514	100.0	514	14	US-10-013-770-2	Sequence 2, Appli
	2	514	100.0	514	14	US-10-045-180A-2	Sequence 2, Appli
	3	393.5	76.6	4415	14	US-10-013-770-1	Sequence 1, Appli
	4	393.5	76.6	4415	14	US-10-045-180A-1	Sequence 1, Appli
	5	227	44.2	542	14	US-10-045-180A-8	Sequence 8, Appli
	6	227	44.2	860	13	US-10-044-090-473	Sequence 473, App
	7	218.0	42.5	555	10	US-10-252-157-223	Sequence 223, App
	8	216.5	42.1	726	15	US-09-816-828-16	Sequence 16, Appl
	9	216.5	42.1	1348	15	US-10-125-237-78	Sequence 78, Appl
	10	216.5	42.1	1348	15	US-10-105-891-78	Sequence 78, Appl
	11	175.5	34.1	496	13	US-10-141-645-11	Sequence 11, Appl
	12	165.5	32.2	500	13	US-10-141-645-14	Sequence 14, Appl
	13	165.5	32.2	500	13	US-10-313-994-13	Sequence 13, Appl
c	14	160	31.1	412	10	US-09-736-457-1047	Sequence 1047, Ap
c	15	160	31.1	412	10	US-09-902-941-1047	Sequence 1047, Ap
c	16	160	31.1	412	13	US-09-849-626-1047	Sequence 1047, Ap
c	17	160	31.1	412	13	US-10-017-872-1047	Sequence 1047, Ap
c	18	160	31.1	412	13	US-10-017-754-1047	Sequence 1047, Ap
	19	159.5	31.0	353	11	US-09-918-995-18323	Sequence 18323, A
c	20	157.5	30.6	337	11	US-09-918-995-18198	Sequence 18198, A
c	21	157.5	30.6	337	11	US-09-918-995-19070	Sequence 19070, A
	22	157.5	30.6	495	13	US-10-141-645-16	Sequence 16, Appl
	23	157.5	30.6	495	13	US-10-313-994-15	Sequence 15, Appl
	24	155	30.2	437	13	US-10-141-645-122	Sequence 122, App
	25	153.5	29.9	372	11	US-09-918-995-18938	Sequence 18938, A
	26	152	29.8	437	13	US-10-141-645-124	Sequence 124, App
	27	152	29.6	437	13	US-10-141-645-120	Sequence 120, App
	28	147	28.6	654	13	US-10-027-632-141529	Sequence 141529,
	29	147	28.6	654	14	US-10-027-632-141529	Sequence 141529,
	30	147	28.6	4295	14	US-10-045-180A-7	Sequence 7, Appli
	31	145.5	28.3	243	13	US-10-313-994-28	Sequence 28, Appl
	32	145	28.2	741	13	US-10-027-632-141528	Sequence 141528,
	33	145	28.2	741	13	US-10-027-632-141530	Sequence 141530,
	34	145	28.2	741	14	US-10-027-632-141528	Sequence 141528,
	35	145	28.2	741	14	US-10-027-632-141530	Sequence 141530,
	36	140.5	27.3	592	13	US-10-027-632-30053	Sequence 30053, A
	37	140.5	27.3	592	13	US-10-027-632-30054	Sequence 30054, A
	38	140.5	27.3	592	14	US-10-027-632-30053	Sequence 30053, A
	39	140.5	27.3	592	14	US-10-027-632-30054	Sequence 30054, A
	40	129.5	25.2	597	9	US-09-864-761-15342	Sequence 15342, A
	41	128.5	25.0	598	13	US-10-027-632-48280	Sequence 48280, A
	42	128.5	25.0	598	13	US-10-027-632-48281	Sequence 48281, A
	43	128.5	25.0	598	13	US-10-027-632-48282	Sequence 48282, A
	44	128.5	25.0	598	13	US-10-027-632-48283	Sequence 48283, A
	45	128.5	25.0	598	13	US-10-027-632-48284	Sequence 48284, A

ALIGNMENTS

RESULT 1
US-10-013-770-2

; Sequence 2, Application US/10013770
; Publication No. US20020115151A1
; GENERAL INFORMATION:
; APPLICANT: GENSET SA
; TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;   COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;   THERAPEUTIC APPLICATIONS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/013,770

;   FILING DATE: 10-Dec-2001
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/486,580
;   FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 453 BASE PAIRS
;   TYPE: NUCLEOTIDE
;   STRANDEDNESS: DOUBLE
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE: cDNA
;   ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
US-10-013-770-2
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Alignment Scores:
Pred. No.:

Score:	1.85e-64	
Percent Similarity:	514.00	Length: 453
Best Local Similarity:	100.00%	Matches: 94
Query Match:	100.00%	Conservative: 0
	100.00%	Mismatches: 0
	14	Indels: 0
DB:		Gaps: 0

US-10-045-180A-3 (1-94) x US-10-013-770-2 (1-453)

```
Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGACCCTCACCCTCCTCTGCCTTTCCTCGTGGTGGCCCTTCAGGCGTGGGCAGAG 111

Qy  21 ProLeuGlnAlaArgAlaAlaHisGluMetProAlaGlnLeuLysGlnProProAlaAspAspGln  40
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 112 CCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGCGGAGGCAGGCCTCCAGCAGAATGACCAG 171

Qy  41 AspValValIleTyrPheSerGlyAspAlaAspSerCysSerLeuGlnValProGlySerThr  60
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 172 GATGTGGTCATTTACTTTTCAGGAGATGCAGACAGCTGCTCTCTTCAGGTTCCAGGCTCAACA 231

Qy  61 LysGlyLeuIleCysHisSerCysArgValLeuTyrCysIlePheGluGlyHisGlyGly  80
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 232 AAGGGCTTGATCTGCCATTGCAGAGTACTAACTGCATTTTTGGAGAACATCTTGGTGGGG 291

Qy  81 ThrCysPheIleLeuGlyLysArgTyrProIleCysCysTyr  94
       ||||||||||||||||||||||||||||||||||||||||
Db 292 ACCTGCTTCATCCTTGGTGAAGGCTACCCAATCTGCTGCTAC 333
```

RESULT 2
US-10-045-180A-2
Sequence 2, Application US/10045180A
Publication No. US20020182703A1
GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Compositi
; TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeutic
; FILE REFERENCE: GEN-100D1
; CURRENT APPLICATION NUMBER: US/10/045,180A
; CURRENT FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 09/486,580
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: PCT/FR98/01864
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: FR 97/10823
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 453
; TYPE: DNA

;   ORGANISM: Homo sapiens
; FEATURE:
;   NAME/KEY:
;   LOCATION: CDS
;   LOCATION: (52)..(336)
;   OTHER INFORMATION:
US-10-045-180A-2
; OTHER INFORMATION: Def-X coding sequence

US-10-045-180A-3 (1-94) x US-10-045-180A-2 (1-453)

Alignment Scores:
Pred. No.:

Score:	1.85e-64	
Percent Similarity:	514.00	Length: 453
Best Local Similarity:	100.00%	Matches: 94
Query Match:	100.00%	Conservative: 0
	100.00%	Mismatches: 0
	14	Indels: 0
DB:		Gaps: 0

```
Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGACCCTCACCCTCCTCTGCCTTTCCTCGTGGTGGCCCTTCAGGCGTGGGCAGAG 111

Qy  21 ProLeuGlnAlaArgAlaAlaHisGluMetProAlaGlnLeuLysGlnProProAlaAspAspGln  40
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 112 CCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGCGGAGGCAGGCCTCCAGCAGAATGACCAG 171

Qy  41 AspValValIleTyrPheSerGlyAspAlaAspSerCysSerLeuGlnValProGlySerThr  60
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 172 GATGTGGTCATTTACTTTTCAGGAGATGCAGACAGCTGCTCTCTTCAGGTTCCAGGCTCAACA 231

Qy  61 LysGlyLeuIleCysHisSerCysArgValLeuTyrCysIlePheGluGlyHisGlyGly  80
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 232 AAGGGCTTGATCTGCCATTGCAGAGTACTAACTGCATTTTTGGAGAACATCTTGGTGGGG 291

Qy  81 ThrCysPheIleLeuGlyLysArgTyrProIleCysCysTyr  94
       ||||||||||||||||||||||||||||||||||||||||
Db 292 ACCTGCTTCATCCTTGGTGAAGGCTACCCAATCTGCTGCTAC 333
```

RESULT 3
US-10-013-770-1
Sequence 1, Application US/10013770
Publication No. US20020115151A1
GENERAL INFORMATION:
; APPLICANT: GENSET SA
; TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAc
;   COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;   THERAPEUTIC APPLICATIONS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/013,770
;   FILING DATE: 10-Dec-2001
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/486,580
;   FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 4415 BASE PAIRS
;   TYPE: NUCLEOTIDE

```
;
;   STRANDEDNESS: DOUBLE
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE: DNA
;   ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY:
;   LOCATION: 1836..1874          Exon 1
;   FEATURE:
;   NAME/KEY:
;   LOCATION: 3394..3577          Exon 2
;   FEATURE:
;   NAME/KEY:
;   LOCATION: 4161..4380          Exon 3
;   FEATURE:
;   NAME/KEY:
;   LOCATION: 3406..3408          start CDS
;   FEATURE:
;   NAME/KEY:
;   LOCATION: 4276..4278          stop CDS
;   FEATURE:
;   NAME/KEY:
;   LOCATION: 4374..4379          polyadenylation site
;   FEATURE:
;
US-10-013-770-1

    SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-013-770-1
DB:

Query Match:          76.56%       Gaps:        1
Best Local Similarity: 76.56%      Indels:      196
Percent Similarity:   32.41%       Mismatches:  0
                      32.41%       Conservative:0
Score:                393.50       Matches:     94
Pred. No.:                         Length:      4415
Alignment Scores:                  8.82e-46

US-10-045-180A-3 (1-94) x US-10-013-770-1 (1-4415)

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaLeuTrpAlaGlu 20
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 3406 ATGAGGACCCTCACCCTCCTCCTCCTCGCCTTTCTCCTGGTGGCCCTTCAGGCTCTGTGGGCAGAG 3465

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db 3466 CCCGTCCAGGCAAGAGCTCATGAGATGCCAGCCCCAGAAGCAGCCTTCCAGCAGATGACCAG 3525

Qy   41 AspValValIleTyrPheSerGlyGlyAspAspSerCysSerLeuGln-------- 55
          ||||||||||||||||||||||||||||||||||||
Db 3526 GATGTGGTCATTTACTTTTCAGGAGGATGACAGCTGCTCTCTTCAGGTTCAGGTGAGAGA 3585

Qy   55 -------- 55
          
Db 3586 TGGCCAGCAATGCAGAGCTACAGACTAGACAGAAGGACACAGGAGGACAGAGGCTCTGGAATTGGAT 3645

Qy   55 -------- 55
          
Db 3646 CTCAGTGGCAGAGTGTCACTTAGGTGGCTATACTTAACATCTCGGGTCCTGGATTTTCTCA 3705

Qy   55 -------- 55
          
Db 3706 TATCTAAATGGAATAGAGACCAAAGAAATCTAGAGAGATTTTTCTTCTCCAAAAACTTG 3765

Qy   55 -------- 55
          
Db 3766 ATTCCAAGATATGACTGTGAAATTCACTAGATTTAAGATATAAGGAGAGATGCTACCTAGTT 3825

Qy   55 -------- 55
          
Db 3826 CCTTCTGGAGCCAGACAAACAAGCTTAAGTATATATAGGAAAAATATTTCACCCTGTCTATAT 3885

Qy   55 -------- 55
          
Db 3886 AGGAGGGTTTTAGAACCTGGAGAGAGCCTAAGAAGAATGTGTTCAGGTGTGTTGTGTGATGGGC 3945

Db 3946 AGGAAATCAGAAAAGTGAAGCAAAGAGAAAGGAAGAAGAGTGGTCTCGAATCCTGTGTGACCAGCACTG 4005

Qy   55 -------- 55
          
Db 4006 CTCTGTGTATTTATTCCTATTGACTGAGATTGTTTGTGCTACGGGCTGTGTTAATACAGCCAA 4065

Qy   55 -------- 55
          
Db 4066 CATCACTCATCAGCCAACATGTGACTTCTCCAAGAATCCCTTACCACCCACTGCTGNAC 4125

Qy   56 ----------------------Val-ProGlySerThrLysGlyValLeuCy 65
                                  |||||||||||||||||||||||||
Db 4126 CCCGTACTACTAGCTCCTCTCTCCTCGGGTCCCCAGGGCTCAACAAAAGGGCTTGAC 4185

Qy   65 sHisCysArgValLeuPheGlyGluHisLeuGlyIleTyrCysPheIleIle 85
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db 4186 CCATTGCAGAGTAACTAATACTGCATTTTTGGGAGAACATCTTGGTGGGAGACCTGCTTCATCCT 4245

Qy   85 uGlyGluArgTyrProIleCysCysGlyTyr 94
          ||||||||||||||||||||||||||||
Db 4246 TGGTGAACGCTACCCAATCTGCTGCTAC 4273
```

RESULT 4
US-10-045-180A-1

```
;   Sequence 1, Application US/10045180A
;   Publication No. US20020182703A1
;   GENERAL INFORMATION:
;   APPLICANT: Bougueleret, Lydie
;   APPLICANT: Chumakov, Ilya
;   TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Composition
;   TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeutic
;   FILE REFERENCE: GEN-1100D1
;   CURRENT APPLICATION NUMBER: US/10/045,180A
;   CURRENT FILING DATE: 2001-10-18
;   PRIOR APPLICATION NUMBER: US 09/486,580
;   PRIOR FILING DATE: 2000-02-25
;   PRIOR APPLICATION NUMBER: PCT/FR98/01864
;   PRIOR FILING DATE: 1998-08-28
;   PRIOR APPLICATION NUMBER: FR 97/10823
;   PRIOR FILING DATE: 1997-08-29
;   NUMBER OF SEQ ID NOS: 12
;   SOFTWARE: Patentin version 3.1
;
;   SEQ ID NO 1
;   LENGTH: 4415
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (85)..(85)
;   OTHER INFORMATION: n = a, c, g, or t.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(4415)
;   OTHER INFORMATION: Def-X genomic sequence
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (143)..(143)
;   OTHER INFORMATION: n = a, c, g, or t.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (670)..(670)
;   OTHER INFORMATION: n = a, c, g, or t.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (970)..(970)
;   OTHER INFORMATION: n = a, c, g, or t.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1111)..(1111)
;   OTHER INFORMATION: n = a, c, g, or t.
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
-;
-;   LOCATION: (1150)..(1150)
-;   OTHER INFORMATION: n = a, c, g, or t.
-;  FEATURE:
-;   NAME/KEY: CAAT signal
-;   LOCATION: (1711)..(1714)
-;   OTHER INFORMATION:
-;  FEATURE:
-;   NAME/KEY: TATA signal
-;   LOCATION: (1758)..(1767)
-;   OTHER INFORMATION:
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (1780)..(1780)
-;   OTHER INFORMATION: n = a, c, g, or t.
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (1836)..(1874)
-;   OTHER INFORMATION: Exon 1
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (1875)..(1880)
-;   OTHER INFORMATION: splice donor site
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (1974)..(1974)
-;   OTHER INFORMATION: n = a, c, g, or t.
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (2117)..(2117)
-;   OTHER INFORMATION: n = a, c, g, or t.
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (2133)..(2133)
-;   OTHER INFORMATION: n = a, c, g, or t.
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (2155)..(2335)
-;   OTHER INFORMATION: Alu insertion
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (2186)..(2186)
-;   OTHER INFORMATION: n = a, c, g, or t.
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (2191)..(2191)
-;   OTHER INFORMATION: n = a, c, g, or t.
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (2367)..(2367)
-;   OTHER INFORMATION: n = a, c, g, or t.
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (2710)..(2780)
-;   OTHER INFORMATION: L1 fragment insertion
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (3391)..(3393)
-;   OTHER INFORMATION: splice acceptor site
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (3394)..(3577)
-;   OTHER INFORMATION: Exon 2
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (3406)..(3408)
-;   OTHER INFORMATION: Translation initiation codon (ATG)
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (3578)..(3583)
-;   OTHER INFORMATION: splice donor site
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (4123)..(4123)
```

```
-;   OTHER INFORMATION: n = a, c, g, or t.
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (4161)..(4163)
-;   OTHER INFORMATION: splice acceptor site
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (4164)..(4379)
-;   OTHER INFORMATION: Exon 3
-;  FEATURE:
-;   NAME/KEY: misc_feature
-;   LOCATION: (4274)..(4276)
-;   OTHER INFORMATION: Translation termination codon (TAA)
-;  FEATURE:
-;   NAME/KEY: polyA_signal
-;   LOCATION: (4374)..(4379)
-;   OTHER INFORMATION:
-;
-; US-10-045-180A-1
```

```
US-10-045-180A-3 (1-94) x US-10-045-180A-1 (1-4415)

DB:
Query Match:        76.56%          Gaps:          1
Best Local Similarity:  32.41%      Indels:        196
Percent Similarity: 32.41%          Mismatches:    0
Score:              393.50          Conservative:  0
Pred. No.:          8.82e-46        Matches:       94
Alignment Scores:                   Length:        4415
```

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3406 ATGAGAACCCTCACCCTCCTCTCTGCCTTTCTCCTGGTTGCCCTTCAGGCCTGGGCAGAG 3465

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3466 CCGCTCCAGGCCAGAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGACCAG 3525

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGln------------   55
          ||||||||||||||||||||||||||||||||||||||
Db 3526 GATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGTGAGGA 3585

Qy   55 -------------------------------------------------------   55
Db 3586 TGCCAGGCATGCAGAGCTACAGACTAGACAGAGGACAGGAGACAGGCTCTGGAATGGAT 3645

Qy   55 -------------------------------------------------------   55
Db 3646 CTCAGTGCAGATGTCACTTAGGTGGCTATACTTAACATCTCTGGTCCTGGATTTTCTCA 3705

Qy   55 -------------------------------------------------------   55
Db 3706 TATCTAAATGGGAATAGAGAACCAAAGAAATCTAAGAGAGATTTTTCTTTCTCCAAAAACTTG 3765

Qy   55 -------------------------------------------------------   55
Db 3766 ATTCCAAGATATGACTGTGAAATTCACTAGATTTAGATATAAGGAGAGATGCTACCTAGTT 3825

Qy   55 -------------------------------------------------------   55
Db 3826 CCTTCTGGAGCCAGGACAAACAAGCTTAAGTATATAGGAAAAAATATATTCACCCTGTCTCTATAT 3885

Qy   55 -------------------------------------------------------   55
Db 3886 AGGAGGTTTTAGAACCTGGAGAGGAGGCCTAAGAAATGTGTTCAGGTGTGTGTGTGATCGGC 3945

Qy   55 -------------------------------------------------------   55
Db 3946 AGGAAATGCAGAAAAGTGAAGCAAAAGAAAGAATCAGTCTCGAATCCCTGTGTGACCAGCACTG 4005

Qy   55 -------------------------------------------------------   55
Db 4006 CTCTGTGTATTTATTCCTTATGACTGAGATTGTGTTTGTGCTACCGGCTGTGTAATACAGCCAA 4065
```

Qy   55 -- 55

Qy   55 -- 55

Db 4066 CATCACTCATCAGCCAAACATGTGACTTCTCCAAGATTCCCTTACCACCACTGCTGNAC 4125

Db 4126 CCCGTACTCAGTTTCTGATGCTCTCTCTCTCCTCGGGTCCCCAGGCTCAACAAAGGCTTGAATCTG 4185
Qy   56 --------------------Val-ProGlySerThrLysGlyLeuIleCy 65

Db 4186 CCATTGCAGAGTACTACTGCCATTTTGGAGAACATCTTGGTGTGGGGACCTGCTTCATCCT 4245
Qy   65 sHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIleLe 85

Db 4246 TGGTGAACGCTACCCAAATCTGCTGCTAC 4273
Qy   85 uGlyGluArgTyrProIleCysCysTyr 94

RESULT 5
US-10-045-180A-8
; Sequence 8, Application US/10045180A
; Publication No. US20020182703A1
; GENERAL INFORMATION:
;   APPLICANT: Bouguelerat, Lydie
;   APPLICANT: Chumakov, Ilya
;   TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Compositi
;   TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeutic
;   FILE REFERENCE: GEN-100D1
;   CURRENT APPLICATION NUMBER: US/10/045,180A
;   CURRENT FILING DATE: 2001-10-18
;   PRIOR APPLICATION NUMBER: US 09/486,580
;   PRIOR APPLICATION NUMBER: PCT/FR98/01864
;   PRIOR FILING DATE: 2000-02-25
;   PRIOR APPLICATION NUMBER: PCT/FR98/01864
;   PRIOR FILING DATE: 1998-08-28
;   PRIOR APPLICATION NUMBER: FR 97/10823
;   PRIOR FILING DATE: 1997-08-29
;   NUMBER OF SEQ ID NOS: 12
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
;   LENGTH: 542
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (52)..(345)
;   OTHER INFORMATION: Def-4 (HNP-4) coding sequence
US-10-045-180A-8 LENGTH: 542

Alignment Scores:
Pred. No.:   4.93e-23
Score: 227.00
Percent Similarity: 64.52%
Best Local Similarity: 49.46%
Query Match: 44.16%

Length: 542
Matches: 46
Conservative: 14
Mismatches: 33
Indels: 0
Gaps: 0

US-10-045-180A-3 (1-94) x US-10-045-180A-8 (1-542)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuLeuValAlaLeuGlnAlaLeuProAlaLeuGlnAlaTrpAlaGlu 20
          ||||||| ::: ||||||||| |||||||||||
Db   1 ATGAGGATTATCGCCCTTCCTGCTATTCTCTTGGTAGCCCTCCAGTCCGGGGCAGGC 52

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
          ||| ||| :::||| :::|||||||
Db   52 ATGAGGATTATCGCCCTTCCTGCTATTCTCTTGGTAGCCCTCCAGTCCGGGGCAGGC 111

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
          |||::: |||||| |||||| :::|||||||
Db   112 CCACTCCAGGCAAGAGGTGATGAGGCCTCCAGGCAGGAGCCAGAGACCAG 171

Qy   61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly 80
          ::|||||||:::||::||| ::: |||
Db   172 GACATATCTATTTCCTTTGCATGGGATAAAAGCTTCGCTTCAGGTTTCAGGTTCAACA 231

Qy   81 ThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
          ||| ::: |||
Db   232 AGGGGCATGGTGTCTCCTTGCAGATTATGTATTCTTGCCGCGAACAGAACTTCGTGTTGGG 291

RESULT 6
US-10-044-090-473
; Sequence 473, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
;   APPLICANT: Olga Bandman
;   TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;   FILE REFERENCE: PA-0028 US
;   CURRENT APPLICATION NUMBER: US/10/044,090
;   CURRENT FILING DATE: 2002-01-09
;   NUMBER OF SEQ ID NOS: 850
;   SOFTWARE: PERL Program
; SEQ ID NO 473
;   LENGTH: 860
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020137081A1 1670142CB1
US-10-044-090-473

Alignment Scores:
Pred. No.:   9.16e-23
Score: 227.00
Percent Similarity: 64.52%
Best Local Similarity: 49.46%
Query Match: 44.16%

Length: 860
Matches: 46
Conservative: 14
Mismatches: 33
Indels: 0
Gaps: 0

US-10-045-180A-3 (1-94) x US-10-044-090-473 (1-860)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuLeuValAlaLeuGlnAlaLeuProAlaLeuGlnAlaTrpAlaGlu 20
          ||||||| ::: ||||||||| |||||||||||
Db   76 ATGAGGATTATCGCCCTTCCTGCTATTCTCTTGGTAGCCCTCCAGTCCGGGGCAGGC 135

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
          ||| ||| :::||| :::|||||||
Db   136 CCACTCCAGGCAAGAGGTGATGAGGCCTCCAGGCAGGAGCCAGAGACCAG 195

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
          |||::: |||||| |||||| :::|||||||
Db   196 GACATATCTATTTCCTTTGCATGGGATAAAAGCTTCGCTTCAGGTTTCAGGTTCAACA 255

Qy   61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly 80
          ::|||||||:::||::||| ::: |||
Db   256 AGGGGCATGGTGTCTCCTTGCAGATTAGTATTCTCCGGCGGAACAGAACTTCGTGTTGGG 315

Qy   81 ThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
          ||| ::: |||
Db   316 AACTGCCTCATTGGTGTGGTGTGAGTTTCACATACTGCTGC 354

RESULT 7
US-10-252-157-223
; Sequence 223, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
;   APPLICANT: Faris, Mary
;   APPLICANT: Pearson, Cecelia I.
;   TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
;   FILE REFERENCE: PA-0027-1 US
;   CURRENT APPLICATION NUMBER: US/10/252,157
;   CURRENT FILING DATE: 2002-10-01
;   PRIOR APPLICATION NUMBER: 60/295,048
;   PRIOR FILING DATE: 2001-05-31
;   NUMBER OF SEQ ID NOS: 501
;   SOFTWARE: PERL Program
; SEQ ID NO 223
;   LENGTH: 555
;   TYPE: DNA
;   ORGANISM: Homo sapiens

Db   292 AACTGCCTCATTGGTGTCGTGTGAGTTTCACATACTGCTGC 330

Qy   81 ThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93

; FEATURE:
;   NAME/KEY: misc feature
;   OTHER INFORMATION: Incyte ID No. US20030190640A1 903338.12
US-10-252-157-223
DB:

Alignment Scores:
Pred. No.:
Score: 8.51e-22
Score: 218.50   Length: 555
Percent Similarity: 64.89%   Matches: 47
Percent Identity: 50.00%   Conservative: 14
Best Local Similarity: 42.51%   Mismatches: 32
Query Match: 13   Indels: 1
DB: Gaps: 1

US-10-045-180A-3 (1-94) x US-10-252-157-223 (1-555)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 145 ATGAGGACCCTCCGCCATCCTTGCTGCCATTCTTCCTGGTTGGCCCTGGGCCCAGGGCTGAG 204

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
        |||||||||||||||| ::
Db 205 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGGACGACGATTGCACGGACATC 264

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
        :||||||||||:::
Db 265 CCAGAGGTGGTTGTTGTTCCCTTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 324

Qy   60 ThrLysGlyLeuLeuCysHisCysArgValLeuTyrCysIlePheGlyHisLeuGly 79
        ::: |||||||||
Db 325 AGGAAAAACATGGCCTCTATTGCAGAATACCAGCCGTGCATTGCAGAGAACGTCGGCTAT 384

Qy   80 GlyThrCysPheIleLeuGlyLeuGluArgTyrProIleCysCys 93
        ::: ||||||
Db 385 GGAACCTGCATCTACCAGGGAGAACTCTGGGCCATTCTGCTGC 426

RESULT 8
US-09-816-828-16
; Sequence 16, Application US/09816828
; Patent No. US20020150898A1
; GENERAL INFORMATION:
;   APPLICANT: Tang, Y. Tom
;   APPLICANT: Zhou, Ping
;   APPLICANT: Goodrich, Ryle
;   APPLICANT: Asundi, Vinod
;   APPLICANT: Ren, Feiyan
;   APPLICANT: Xue, Aidong J.
;   APPLICANT: Ma, Yunqing
;   APPLICANT: Wang, Zhiwei
;   APPLICANT: Zhao, Qing A.
;   APPLICANT: Zhang, Jie
;   APPLICANT: Wang, Jian-Rui
;   APPLICANT: Drmanac, Radoje T.
;   TITLE OF INVENTION: Polypeptides
;   FILE REFERENCE: 791CIP2B
;   CURRENT APPLICATION NUMBER: US/09/816,828
;   CURRENT FILING DATE: 2001-03-22
;   PRIOR APPLICATION NUMBER: 09/770,160
;   PRIOR FILING DATE: 2001-01-26
;   PRIOR APPLICATION NUMBER: 09/552,929
;   PRIOR FILING DATE: 2000-04-18
;   NUMBER OF SEQ ID NOS: 18
;   SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 16
;   LENGTH: 726
;   TYPE: DNA
;   ORGANISM: Homo sapiens
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (151)..(522)
US-09-816-828-16

Alignment Scores:

Pred. No.:
Score: 2.37e-21
Score: 216.50   Length: 726
Percent Similarity: 64.89%   Matches: 47
Percent Identity: 50.00%   Conservative: 14
Best Local Similarity: 42.12%   Mismatches: 32
Query Match: 10   Indels: 1
DB: Gaps: 1

US-10-045-180A-3 (1-94) x US-09-816-828-16 (1-726)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 ATGAGGACCCTCCGCCATCCTTGCTGCCATTCTTCCTGGTTGGCCCTGGGCCCAGGGCTGAG 300

Qy   21 ProLeuGlnAlaArgAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
        |||||||||||||||| ::
Db 301 CCACTCCAGGCAAGAGCTGATGAGGTTGCTGCAGCCCCGGACGACGATTGCACGGACATC 360

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
        :||||||||||:::
Db 361 CCAGAGGTGGTTGTTGTTTCCCTTGCATGGGACGAAAGCTTGGCTCCAAAGCATCCAGGCTCA 420

Qy   60 ThrLysGlyLeuLeuCysHisCysArgValLeuTyrCysIlePheGlyHisLeuGly 79
        ::: |||||||||
Db 421 AGGAAAAACATGGCCTCTATTGCAGAATACCAGCCGTGCATTGCAGAGAACGTCGGCTAT 480

Qy   80 GlyThrCysPheIleLeuGlyLeuGluArgTyrProIleCysCys 93
        ::: ||||||
Db 481 GGAACCTGCATCTACCAGGGAGAACTCTGGGCCATTCTGCTGC 522

RESULT 9
US-10-125-237-78
; Sequence 78, Application US/10125237
; Publication No. US20030223291A1
; GENERAL INFORMATION:
;   APPLICANT: Tang, Y. Tom
;   APPLICANT: Liu, Chenghua
;   APPLICANT: Zhou, Ping
;   APPLICANT: Asundi, Vinod
;   APPLICANT: Ren, Feiyan
;   APPLICANT: Zhao, Qing A.
;   APPLICANT: Xue, Aidong J.
;   APPLICANT: Zhang, Jie
;   APPLICANT: Wehrman, Tom
;   APPLICANT: Wang, Jian-Rui
;   APPLICANT: Drmanac, Radoje T.
;   TITLE OF INVENTION: No. US20030223291A1e1 Nucleic Acids and
;   TITLE OF INVENTION: Polypeptides
;   FILE REFERENCE: 791CIP2ADIV
;   CURRENT APPLICATION NUMBER: US/10/125,237
;   CURRENT FILING DATE: 2002-04-17
;   PRIOR APPLICATION NUMBER: 09/668,317
;   PRIOR FILING DATE: 2000-09-22
;   PRIOR APPLICATION NUMBER: 09/552,929
;   PRIOR FILING DATE: 2000-04-18
;   NUMBER OF SEQ ID NOS: 91
;   SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 78
;   LENGTH: 1348
;   TYPE: DNA
;   ORGANISM: Homo sapiens
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1143)
US-10-125-237-78

Alignment Scores:
Pred. No.:
Score: 5.45e-21
Score: 216.50   Length: 1348
Percent Similarity: 64.89%   Matches: 47
Percent Identity: 50.00%   Conservative: 14
Best Local Similarity: 42.12%   Mismatches: 32
Query Match: 15   Indels: 1
DB: Gaps: 1

US-10-045-180A-3 (1-94) x US-10-125-237-78 (1-1348)

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
        ||||||||||||| ::||:|||||||||||
Db  859 ATGAGGACCCTGCANCCTTGCNCCATTCTCCTGGTGGCCCTGCAGGCCCCAGGCTGA 918

Qy   21 ProLeuGlnAlaArgAlaAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
        |||||||||||| :: |||||||||||
Db  919 CCACTCCAGGGCAAGAGCTGTGAGGTTGCTGCAGCCCCGGGACAGATTGCAGCGGGACATC 978

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
        |||:|||||||
Db  979 CCAGAAGTGGTTGTTTCCCTTGCATGGGACGACAGAAGCTTGGCCCAGAGATGCACGGGGACATC 1038

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
        :|| ||:||||:|:||||
Db 1039 AGGAAAACATGGACTGCTATTGCGAGAATACCAGCGTGCATTGCAGGAACGTCGCTAT 1098

Qy   80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
        :|| |||
Db 1099 GGAACCTGCATCTACCAGGGAAGAGACTCTGGGCATTCTGCTGC 1140
```

RESULT 10
US-10-105-891-78
US-10-105-891-78

Alignment Scores:

Pred. No.:		
Score:	216.50	Matches: 47
Percent Similarity:	64.89%	Conservative: 14
Best Local Similarity:	50.00%	Mismatches: 32
Query Match:	42.12%	Indels: 1
DB:	Length: 1348	Gaps: 1

US-10-045-180A-3 (1-94) x US-10-105-891-78 (1-1348)

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
        ||||||||||||| ::||:||||||||||
Db    1 ATGAGGACCCTCGCCATTCTTCTGCTGGTGCCCCTGCAGGCCCCACGGCTGAG 20

Qy   21 ProLeuGlnAlaArgAlaAlaHisGluMet---ProAlaGlnLysGlnProProAlaAspAsp 39
        |||||||||||:|||:|||||||
Db  859 ATGAGGACCCTCGCCCATTCTCCTGGTGCCCCTGCAGGCCCCACGGCTGAG 918
```

```
Db  919 ||||||||||||| :::
Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
        ::|||||||||||
Db  979 CCAGAAGTGGTTGTTTCCCTTGCATGGGACGACAGAAGCTTGGCCCAGAGATGCACGGGGACATC 1038

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
        :|| ||:||||:|:||||
Db 1039 AGGAAAACATGGACTGCTATTGCGAGAATACCAGCGTGCATTGCAGGAACGTCGCTAT 1098

Qy   80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
        :|| |||
Db 1099 GGAACCTGCATCTACCAGGGAAGAGACTCTGGGCATTCTGCTGC 1140
```

RESULT 11
US-10-141-645-11
US-10-141-645-11

Alignment Scores:

Pred. No.:		
Score:	1.14e-15	Matches: 44
Percent Similarity:	175.50	Conservative: 11
Best Local Similarity:	58.51%	Mismatches: 38
Query Match:	46.81% / 34.14%	Indels: 1
DB:	Length: 496	Gaps: 1

US-10-045-180A-3 (1-94) x US-10-141-645-11 (1-496)

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaLeuTrpAla 19
        ||||||||||||| ::||:||||||||||
Db   76 ATGAGGACCCTCGCCCTCCTCACTGCCTTCTCCTGGTGGCCCTGCAGGCG 135

Qy   20 GluProLeuGlnAlaArgAlaAlaHisGluMetProAlaGlnLysGlnProProAlaAsp 39
        |||||||||::|||:|||||||
Db  136 GAGCCACTTCAGGCAAGAGCTGCAGCCCAGAAGCAGCCGGGAGATGAT 195

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
        ||:||||||
Db  196 CAGGAAATGGGCTCATGCCTTACATGGCATGGAAAGTGCCGTCTTCCGCTTCA 255

Qy   60 ThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGly 79
        |||:||:|||:|:||||
Db  256 GCGAGAGGCTTGAGGTGCATTGCGAAGAGGAATTGCGTTTGTTATAACGTCGCTTT 315

Qy   80 GlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
        :|| |||
```

Db
316 GGGTCCTGCCGCCTTTCGTGGTACACTCCACCGGAATCTGCTGC 357

RESULT 12
US-10-141-645-14
; Sequence 14, Application US/10141645
; Publication No US20030141184A1
; GENERAL INFORMATION:
; APPLICANT: Robert Lehrer
; APPLICANT: Alan Waring
; APPLICANT: Alexander Cole
; APPLICANT: Teresa Hong
; TITLE OF INVENTION: Retrocyclins - Antiviral and
; TITLE OF INVENTION: Antimicrobial Peptides
; FILE REFERENCE: UCLA-001CIP
; CURRENT APPLICATION NUMBER: US/10/141,645
; CURRENT FILING DATE: 2002-05-06
; PRIOR APPLICATION NUMBER: 60/284,855
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: Unassigned
; PRIOR FILING DATE: 2002-04-18
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 14
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Macaca mulatta
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (95)...(325)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (95)...(154)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (287)...(313)
; OTHER INFORMATION: ligated to RTD1b in head-to-tail orientation to
; OTHER INFORMATION: form the cyclic octadecapeptide RTD1; RTD1 is
; OTHER INFORMATION: stabilized by three intramolecular disulfides

US-10-141-645-14

Alignment Scores:
Pred. No.:
Score: 3.16e-14
                          165.50
Percent Similarity: 56.57%
Best Local Similarity: 45.45%
Query Match: 32.20%
DB:

Length: 500
Matches: 45
Conservative: 11
Mismatches: 32
Indels: 11
Gaps: 4

US-10-045-180A-3 (1-94) x US-10-141-645-14 (1-500)

Qy   1 MetArgThrLeuThrLeuLeuSerAla----PheLeuLeuValAlaLeuGlnAlaTrpAla 19
          ||||||||||||||||||||| ::|||
Db   95 ATGAGGACCTTGCGCCCTCCTCACCGACCATGCTTCTCCTGGTGGGCCCTGCACGGCTCAGGCA 154

Qy   20 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 39
          ||||||||||||||||||||||| :: |||
Db 155 GAGGCACGTCAGGCAAGAGCTGCCCGCCCAGCAGCCTGGAACAGATGAT 214

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
          ||||| :::   ::|||
Db 215 CAGGGAAATGGCCTCATTCCTTTACATGGCCTGAAAACGCCCTCTTCCACTTTCAGAGTCA 274

Qy   60 ThrLysGlyLeuIleCys----------HisCysArgValLeuTyrCysIlePhe 74
          ||| :::|||||||
Db 275 GCCGAAAGGCTTGAGGTGCATTTGCACACGAGGAGAATTCTGCCGTTTGTTATAAGT------ 328

Qy   75 GlyGluHisLeuGlyGlyThrCysPheIleLeuLeuGlyLeuArgTyrProIleCysCys 93
          |||| :::|||
Db 329 ------CACCTT--GGGTCCTGCCGCCTTTCGTGGTACACTCCACCGGAATCTGCTGC 376

RESULT 13
US-10-313-994-13

; Sequence 13, Application US/10313994
; Publication No. US20030162718A1
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Tang, Yi-Quan
; APPLICANT: Yuan, Jun
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antimicrobial Theta Defensins and Methods of Using Same
; FILE REFERENCE: P-UC 3095
; CURRENT APPLICATION NUMBER: US/10/313,994
; CURRENT FILING DATE: 2002-12-05
; PRIOR APPLICATION NUMBER: US/09/309,487
; PRIOR FILING DATE: 1999-05-10
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 13
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Macaca mulatta
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (95)...(325)
; US-10-313-994-13

Alignment Scores:
Pred. No.:
Score: 3.16e-14
                          165.50
Percent Similarity: 56.57%
Best Local Similarity: 45.45%
Query Match: 32.20%
DB:

Length: 500
Matches: 45
Conservative: 11
Mismatches: 32
Indels: 11
Gaps: 4

US-10-045-180A-3 (1-94) x US-10-313-994-13 (1-500)

Qy   1 MetArgThrLeuThrLeuLeuSerAla----PheLeuLeuValAlaLeuGlnAlaTrpAla 19
          ||||||||||||||||||||| ::|||
Db   95 ATGAGGACCTTGCGCCCTCCTCACCGACCATGCTTCTCCTGGTGGGCCCTGCACGGCTCAGGCA 154

Qy   20 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 39
          ||||||||||||||||||||||| :: |||
Db 155 GAGGCACGTCAGGCAAGAGCTGCCCGCCCAGCAGCCTGGAACAGATGAT 214

Qy   40 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 59
          ||||| :::   ::|||
Db 215 CAGGGAAATGGCCTCATTCCTTTACATGGCCTGAAAACGCCCTCTTCCACTTTCAGAGTCA 274

Qy   60 ThrLysGlyLeuIleCys----------HisCysArgValLeuTyrCysIlePhe 74
          ||| :::|||||||
Db 275 GCCGAAAGGCTTGAGGTGCATTTGCACACGAGGAGAATTCTGCCGTTTGTTATAAGT------ 328

Qy   75 GlyGluHisLeuGlyGlyThrCysPheIleLeuLeuGlyLeuArgTyrProIleCysCys 93
          |||| :::|||
Db 329 ------CACCTT--GGGTCCTGCCGCCTTTCGTGGTACACTCCACCGGAATCTGCTGC 376

RESULT 14
US-09-736-457-1047/c
; Sequence 1047, Application US/09736457
; Patent No. US20020168837A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457

```
;  CURRENT FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 1864
;  SOFTWARE: FastSEQ for Windows Version 3.0
;
;  SEQ ID NO 1047
;    LENGTH: 412
;    TYPE: DNA
;    ORGANISM: Homo sapien
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)..(412)
;    OTHER INFORMATION: n = A,T,C or G

US-09-736-457-1047

Alignment Scores:
Pred. No.:                         1.51e-13
Score:                             160.00
Percent Similarity:                53.09%
Best Local Similarity:             43.21%
Query Match:                       31.13%
DB:                                10

                       Length:     412
                       Matches:    35
                       Conservative: 8
                       Mismatches: 38
                       Indels:     0
                       Gaps:       0

US-10-045-180A-3 (1-94) x US-09-736-457-1047 (1-412)

Qy      13 ValAlaLeuGlnAlaTrpAlaGluProLeuGlnAlaArgAlaHisGluMetProAlaGln 32
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     410 GTGGGCCCTGCAGGCCCAGGCTGAGTCACTCCAGGAAAGAGCTGACTACAACCAG 351

Qy      33 LysGlnProProAlaAspAspGlnAspValValIleTyrPheSerGlyAspAspSerCys 52
           |||||||||        |||||||:::  |||||||||:::
Db     350 AAGCAGTNTGGGGAAGACAACCAGGACCTTGCTATCTCCTTTGCAGGAAATGGACTNTNT 291

Qy      53 SerLeuGlnValProGlySerThrLysGlyLeuIleCysHisCysArgValLeuTyrCys 72
           ::: :::  |||
Db     290 GCTNTTAGAACCTCAGGTTNTCAGGCAAGAGAGCCACCTGCTATTGCCGAACCGGCCGTTGT 231

Qy      73 IlePheGlyGluHisLeuGlyValGlyThrCysPheIleLeuGlyGluArgTyrProIleCys 92
           |||
Db     230 GCTACCCGTGAGTCCCTCTCCGGGGGTGTGTGAAATCAGTGGCCGGCCTNTACAGACTCTGC 171

Qy      93 Cys 93
           |||
Db     170 TGT 168


RESULT 15
US-09-902-941-1047/c
;  Sequence 1047, Application US/09902941
;  Patent No. US20020172952A1
;  GENERAL INFORMATION:
;    APPLICANT: Henderson, Robert A.
;    APPLICANT: Wang, Tongtong
;    APPLICANT: Watanabe, Yoshihiro
;    APPLICANT: Johnson, Jeffrey C.
;    APPLICANT: Retter, Marc W.
;    APPLICANT: Marmerakis, Margarita
;    APPLICANT: Carter, Darrick
;    APPLICANT: Fanger, Gary R.
;    APPLICANT: Vedvick, Thomas S.
;    APPLICANT: Bangur, Chaitanya S.
;    APPLICANT: McNabb, Andria
;    TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;    TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
;    FILE REFERENCE: 210121.478C17
;    CURRENT APPLICATION NUMBER: US/09/902,941
;    CURRENT FILING DATE: 2001-07-10
;    NUMBER OF SEQ ID NOS: 2002
;    SOFTWARE: FastSEQ for Windows Version 4.0
;
;  SEQ ID NO 1047
;    LENGTH: 412
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: 183, 271, 287, 292, 294, 343
```

```
;    OTHER INFORMATION: n = A,T,C or G

US-09-902-941-1047

Alignment Scores:
Pred. No.:                         1.51e-13
Score:                             160.00
Percent Similarity:                53.09%
Best Local Similarity:             43.21%
Query Match:                       31.13%
DB:                                10

                       Length:     412
                       Matches:    35
                       Conservative: 8
                       Mismatches: 38
                       Indels:     0
                       Gaps:       0

US-10-045-180A-3 (1-94) x US-09-902-941-1047 (1-412)

Qy      13 ValAlaLeuGlnAlaTrpAlaGluProLeuGlnAlaArgAlaHisGluMetProAlaGln 32
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     410 GTGGGCCCTGCAGGCCCAGGCTGAGTCACTCCAGGAAAGAGCTGACTACAACCAG 351

Qy      33 LysGlnProProAlaAspAspGlnAspValValIleTyrPheSerGlyAspAspSerCys 52
           |||||||||        |||||||:::  |||||||||:::
Db     350 AAGCAGTNTGGGGAAGACAACCAGGACCTTGCTATCTCCTTTGCAGGAAATGGACTNTNT 291

Qy      53 SerLeuGlnValProGlySerThrLysGlyLeuIleCysHisCysArgValLeuTyrCys 72
           ::: :::  |||
Db     290 GCTNTTAGAACCTCAGGTTNTCAGGCAAGAGAGCCACCTGCTATTGCCGAACCGGCCGTTGT 231

Qy      73 IlePheGlyGluHisLeuGlyValGlyThrCysPheIleLeuGlyGluArgTyrProIleCys 92
           |||
Db     230 GCTACCCGTGAGTCCCTCTCCGGGGGTGTGTGAAATCAGTGGCCGGCCTNTACAGACTCTGC 171

Qy      93 Cys 93
           |||
Db     170 TGT 168


Search completed: December 17, 2003, 21:35:57
Job time : 330 secs
```

```
OM protein - nucleic search, using frame_plus_p2n model

                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:             December 17, 2003, 14:50:33  ; Search time 94 Seconds
                                                 (without alignments)
                                                 441.383 Million cell updates/sec

Title:              US-10-045-180A-3
Perfect score:      514
Sequence:           1 MRITLTLLSAFLIVALQAWAE...........GEHLGGTCPILGERYPICCY 94

Scoring table:      BLOSUM62
                    Xgapop 10.0 , Xgapext    0.5
                    Xgapop 10.0 , Ygapext    0.5
                    Fgapop  6.0 , Fgapext    7.0
                    Delop   6.0 , Delext     7.0

Searched:           569978 seqs, 220691566 residues

Post-processing:    Minimum Match 0%
                    Maximum Match 100%
                    Listing first 45 summaries

Total number of hits satisfying chosen parameters:    1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
-MODEL=frame+_p2n.model  -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145024_12553/app_query.fasta_1.860
-DB=Issued_Patents_NA  -QFMT=fastap  -SUFFIX=rni  -MINMATCH=0.1  -LOOPCL=0
-LIST=45  -DOCALIGN=200  -NORM=ext  -HEAPSIZE=500  -MINLEN=0  -MAXLEN=200000000
-MODE=LOCAL  -OUTFMT=pct  -THR_SCORE=pct  -THR_MIN=0  -ALIGN=15
-USER=US10045180 @CGN_1 1_193 @runat 17122003_145024_12553  -NCPU=6  -ICPU=3
-NO_MMAP  -LARGEQUERY  -NEG_SCORES=0  -WAIT  -DSPBLOCK=100  -LONGLOG
-DEV_TIMEOUT=120  -WARN_TIMEOUT=30  -THREADS=1  -XGAPOP=10  -XGAPEXT=0.5  -FGAPOP=6
-FGAPEXT=7  -YGAPOP=10  -YGAPEXT=0.5  -DELOP=6  -DELEXT=7
-LOOPEXT=0  -UNITS=Bits  -START=1  -END=-1  -MATRIX=blosum62  -TRANS=human40.cdi
-DB=Issued_Patents_NA  -QFMT=fastap  -SUFFIX=rni  -MINMATCH=0.1  -LOOPCL=0

Database :          Issued_Patents_NA:*
          1: /cgn2_6/ptcdata/2/ina/5A_COMB.seq:*
          2: /cgn2_6/ptcdata/2/ina/5B_COMB.seq:*
          3: /cgn2_6/ptcdata/2/ina/6A_COMB.seq:*
          4: /cgn2_6/ptcdata/2/ina/6B_COMB.seq:*
          5: /cgn2_6/ptcdata/2/ina/PCTUS_COMB.seq:*
          6: /cgn2_6/ptcdata/2/ina/backfiles1.seq:*


Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	514	100.0	453	4	US-09-486-580A-2	Sequence 2, Appli
2	393.5	76.6	4415	4	US-09-486-580A-1	Sequence 1, Appli
3	220	42.8	452	1	US-08-158-189-6	Sequence 6, Appli
4	220	42.8	452	4	US-09-016-434-1198	Sequence 1198, Ap
5	207	40.3	424	4	US-08-158-189-4	Sequence 4, Appli
6	173.5	33.8	420	1	US-08-486-013-46	Sequence 46, Appli
7	173.5	33.8	420	2	US-08-482-279-46	Sequence 46, Appl
8	173.5	33.8	420	3	US-08-342-268-46	Sequence 46, Appl
9	173.5	33.8	420	4	US-09-015-968-46	Sequence 46, Appl
10	173.5	33.8	420	4	US-09-397-386-46	Sequence 46, Appl
11	171.5	33.4	401	1	US-08-486-013-54	Sequence 54, Appl
12	171.5	33.4	401	2	US-08-482-279-54	Sequence 54, Appl
13	171.5	33.4	401	2	US-08-342-268-54	Sequence 54, Appl
14	171.5	33.4	401	4	US-09-015-968-54	Sequence 54, Appl
15	168.5	32.8	401	4	US-09-397-386-54	Sequence 54, Appl
16	168.5	32.8	388	1	US-08-486-013-52	Sequence 52, Appl
17	168.5	32.8	388	2	US-08-482-279-52	Sequence 52, Appl
18	168.5	32.8	388	2	US-08-342-268-52	Sequence 52, Appl
19	168.5	32.8	388	3	US-09-015-968-52	Sequence 52, Appl
20	168.5	32.8	388	4	US-09-397-386-52	Sequence 52, Appl
21	166.5	32.4	377	1	US-08-486-013-48	Sequence 48, Appl
22	166.5	32.4	377	2	US-08-482-279-48	Sequence 48, Appl
23	166.5	32.4	377	2	US-08-342-268-48	Sequence 48, Appl
24	166.5	32.4	377	3	US-09-015-968-48	Sequence 48, Appl
25	166.5	32.4	377	4	US-09-397-386-48	Sequence 48, Appl
26	165.5	32.2	422	1	US-08-342-268-41	Sequence 41, Appl
27	165.5	32.2	422	2	US-08-482-279-41	Sequence 41, Appl
28	165.5	32.2	422	2	US-08-486-013-41	Sequence 41, Appl
29	165.5	32.2	422	4	US-09-397-386-41	Sequence 41, Appl
30	165.5	32.2	422	4	US-09-015-968-41	Sequence 41, Appl
31	165.5	32.2	500	4	US-09-487-13	Sequence 13, Appl
32	165.5	32.2	500	4	US-09-967-808-13	Sequence 13, Appl
33	164.5	32.0	422	1	US-08-486-013-42	Sequence 42, Appl
34	164.5	32.0	422	2	US-08-482-279-42	Sequence 42, Appl
35	164.5	32.0	422	2	US-08-342-268-42	Sequence 42, Appl
36	164.5	32.0	422	3	US-09-015-968-42	Sequence 42, Appl
37	164.5	32.0	422	4	US-09-397-386-42	Sequence 42, Appl
38	163.5	31.8	422	2	US-08-486-013-40	Sequence 40, Appl
39	163.5	31.8	422	2	US-08-482-279-40	Sequence 40, Appl
40	163.5	31.8	422	2	US-08-342-268-40	Sequence 40, Appl
41	163.5	31.8	422	4	US-09-967-808-13	Sequence 13, Appl
42	163.5	31.8	422	4	US-09-001-968-40	Sequence 40, Appl
43	163.5	31.8	422	4	US-09-397-386-40	Sequence 40, Appl
44	160	31.1	365	1	US-08-486-013-43	Sequence 43, Appl
45	160	31.1	365	2	US-08-342-268-43	Sequence 43, Appl

ALIGNMENTS

```
RESULT 1
US-09-486-580A-2
; Sequence 2, Application US/09486580A
; Patent No. 6329340
; GENERAL INFORMATION:
; APPLICANT: GENSET SA
; TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
; TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
; TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 550 West C Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/486,580A
; CURRENT APPLICATION DATA:
; FILING DATE: FEBRUARY 25, 2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Hart, Daniel
; REGISTRATION NUMBER: 40,637
; REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 BASE PAIRS
; TYPE: NUCLEOTIDE
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: cDNA
```

; ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
US-09-486-580A-2
;
Alignment Scores:
Pred. No.:
Score: 1.06e-56
Percent Similarity: 514.00   Length: 453
Best Local Similarity: 100.00%   Matches: 94
Query Match: 100.00%   Conservative: 0
DB: 4   Mismatches: 0
                                       Indels: 0
                                       Gaps: 0

US-10-045-180A-3 (1-94) x US-09-486-580A-2 (1-453)

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   52 ATGAGGACCCTCACCCTCCTCTCTGCCTTTCTCCTGGTGGCCCTTCAGGCCCTGGGCAGAG 111

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspAspGln 40
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  112 CCGCTCCAGGCAAGGAGCTCATGAGATGCCCAGCCCAGAAGCAGCCTCCAGCAGATGACCAG 171

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  172 GATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGGTCAACA 231

Qy   61 LysGlyLeuIleLeuCysHisCysArgValLeuTyrCysIlePheGlyGluHisIleLeuGlyGly 80
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  232 AAGGGCTTGAATCTGCCATTGCAGAGTACTATACTGCCATTTTTGGAGAACATCTTGGGTGGG 291

Qy   81 ThrCysPheIleLeuGlyGluArgThrTyrProIleCysCysTyr 94
         ||||||||||||||||||||||||||||||||||||
Db  292 ACCTGCTTCATCCTTGGTGAACGGCTACCCAAATCTGCTGCTAC 333
```

RESULT 2
US-09-486-580A-1
Sequence 1, Application US/09486580A
Patent No. 6329940
GENERAL INFORMATION:
  APPLICANT: GENSET SA
  TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
  TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
  TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
  NUMBER OF SEQUENCES: 6
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Knobbe, Martens, Olson & Bear
    STREET: 550 West C Street
    CITY: San Diego
    STATE: California
    COUNTRY: USA
    ZIP: 92101
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/486,580A
    FILING DATE: FEBRUARY 25, 2000
  ATTORNEY/AGENT INFORMATION:
    NAME: Hart, Daniel
    REGISTRATION NUMBER: 40,637
    REFERENCE/DOCKET NUMBER: GENSET.064C1
  INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 4415 BASE PAIRS
      TYPE: NUCLEOTIDE
      STRANDEDNESS: DOUBLE
      TOPOLOGY: LINEAR
    MOLECULE TYPE: DNA
    ORIGINAL SOURCE:
      ORGANISM: Homo sapiens
    FEATURE:

;   NAME/KEY: Exon 1
;   LOCATION: 1836..1874
; FEATURE:
;   NAME/KEY: Exon 2
;   LOCATION: 3394..3577
; FEATURE:
;   NAME/KEY: Exon 3
;   LOCATION: 4161..4380
; FEATURE:
;   NAME/KEY: stop CDS
;   LOCATION: 4276..4278
; FEATURE:
;   NAME/KEY: start CDS
;   LOCATION: 3406..3408
; FEATURE:
;   NAME/KEY: polyadenylation site
;   LOCATION: 4374..4379
US-09-486-580A-1

Alignment Scores:
Pred. No.:
Score: 4.72e-40
Percent Similarity: 393.50   Length: 4415
Best Local Similarity: 32.41%   Matches: 94
Query Match: 32.41%   Conservative: 0
DB: 4   Mismatches: 0
                          76.56%   Indels: 196
                                       Gaps: 1

US-10-045-180A-3 (1-94) x US-09-486-580A-1 (1-4415)

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db 3406 ATGAGGACCCTCACCCTCCTCTCTGCCTTTCTCCTGGTGGCCCTTCAGGCCCTGGGCAGAG 3465

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspAspGln 40
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db 3466 CCGCTCCAGGCAAGGAGCTCATGAGATGCCCAGCCCAGAAGCAGCCTCCAGCAGATGACCAG 3525

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGln------ 55
         |||||||||||||||||||||||||||||||||||
Db 3526 GATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGGTCAACAGAGA 3585

Qy   55 ------
Db 3586 TGCCAGCATGCAGAGCTACAGACTAGACAGAAGGACACAGGAGACAGAGGCTCTGGAATTGGAT 3645

Qy   55 ------
Db 3646 CTCAGTGGCAGAGTGCACTTAGGTGGCTATACTTAACATCTCTGGTCCTGGATTTCTCA 3705

Qy   55 ------
Db 3706 TATCTAAATGGAATAGAGAACCAAAGAAAATCTAAGAGAGATTTTTCTTCTCCAAAAACTTG 3765

Qy   55 ------
Db 3766 ATTCCAAGATATGACTGTGAAATCACTAGAATTAAGATAATAAGGAGATGCTACCTAGTT 3825

Qy   55 ------
Db 3826 CCTTCTGGAGCCAGACAAACAAGCTTAAGTATATAGGAAAAATATTTCACCCTGTCTATAT 3885

Qy   55 ------
Db 3886 AGGAGGTTTAGAACCTGGAGGAGGCCTAAGAAGTGTTCAGGTGTGTGTTGATGGGC 3945

Qy   55 ------
Db 3946 AGGAATGCAGAAAAAGTGAAGCAAAGGAAATGAGTCTCGAATCCTGTGTGTGACCAGCACTG 4005

Qy   55 ------
Db 4006 CTCTGTGTATTTATTCCTATTGACTGAGATTGTTTGTGCTACCGGCTGTAATACAGCCAA 4065
```

RESULT 3
US-08-158-189-6
; Sequence 6, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
; APPLICANT: Bevins, Charles L.
; APPLICANT: Jones, Douglas E.
; TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,189
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,232
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Johnson, Philip S.
; REGISTRATION NUMBER: 27,200
; REFERENCE/DOCKET NUMBER: CH-0219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 452 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..321
;
US-08-158-189-6

Alignment Scores:
Pred. No.:
Score: 2.71e-19
Score: 220.00   Length: 452
Percent Similarity: 63.00%   Matches: 49
Best Local Similarity: 49.00%   Conservative: 14
Query Match: 42.80%   Mismatches: 29
DB: Indels: 8
                                      Gaps: 2

Qy   55 --- 55
Db   4066 CATCACTCATCAGCCAACAATGTGACTTCTCCAAGATTCCCTTTACCACCACTGTCTGNAC 4125

Qy   56 ------Val-ProGlySerThrLysGlyLeuIleCy 65
Db   4126 CCCGTACTCAGTTTCTGATGCTCTTCTCTGGGGTCCCAGGCTCAACAAAGGGCTTGATCTG 4185

Qy   65 sHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIleLe 85
Db   4186 CCATTGCAGAGAGTACTATACTGCATTTTTGGAGAACATCTTGGTGGGGACCTGCTTCATCCT 4245

Qy   85 uGlyGlyLeuArgTyrTyrProIleCysCysTyr 94
Db   4246 TGGTGAACGGCTACCCAATCTGCTGCTAC 4273

US-10-045-180A-3 (1-94) x US-08-158-189-6 (1-452)

RESULT 4
US-09-016-434-1198
; Sequence 1198, Application US/09016434
; Patent No. 6509938
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1198:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 452 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 9181546
;
US-09-016-434-1198

Alignment Scores:
Pred. No.:
Score: 2.71e-19
Score: 220.00   Length: 452
Percent Similarity: 63.00%   Matches: 49
Best Local Similarity: 49.00%   Conservative: 14
                                      Mismatches: 29

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuLeuValAlaLeuGlnAlaThrProAlaGlu 20
Db   19 ATGAGAACCCTCACCCTCACTGTCGTTTCCTCGTTCCTGCTGCTGGTGGCCCTGCAAGGCTGAG 78

Qy   21 ------ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLys 33
Db   79 CCACTCCAAGCTCGAGGATGATCCACTCGCAGGCGAAAAGCTTATGAGGGTGCCCAGGAG 138

Qy   34 GlnProProAlaAspAspGlnAspValValIleTyrPheSerGlyAspAspSerCysSer 53
Db   139 CAGCCGGTGGGGCAAAATGACCAGGACTTTGGCGTCTCCTTTGCAGAGGATGCAAGTCAAGT 198

Qy   54 LeuGlnValProGlySerThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIle 73
Db   199 CTTAGAGCTTTGGCTCAACGAAGGGCTTTACTTGCATTGCAGA--AGGTCCTGTTAT 255

Qy   74 PheGlyGluHisLeuGlyGlyThrCysPheIleLeuGlyGlyLeuArgTyrTyrProIleCysCys 93
Db   256 TCAACAGAATATTCCTATGGGACCTGCACTGTCATGGGTGTTATTAACCACAGATTCTGCTGC 315

Query Match: 42.80%   Indels: 8
DB: 4   Gaps: 2

US-10-045-180A-3 (1-94) x US-09-016-434-1198 (1-452)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
          |||:||:||||||:|::||:|
Db   19 ATGAGAACCCTCACCATCCTTCACTGCTGTTCTCTGCTGGCCCTTCCAGGCCTGAG 19

Qy   21 ---------------------ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLys 33
                                |||||||:|:|||:::
Db   79 CCACTCCAAGCTGAGGATGGAGGATGACCATTCCACTCCCAAGAAAGCTTATGAGGGTCAGGAG 138

Qy   34 GlnProProAlaAspAlaAspGlnAspValValIleTyrPheSerGlyAlaAspAspAspProGln 53
          |||:|||:|||||||||:::
Db   139 CAGCCGTGGGGCGCAAATGACCAGGACTTTGCCGTCTCCTTTGCAGAGGATGCCAAGCTCAAGT 198

Qy   54 LeuGlnValProGlySerThrLysGlyLeuIleCysHisCysArgValLeuTyrCysIle 73
          |||:|||:::||:
Db   199 CTTAGAGCTTTGGGCCTCAAACAAGGGCTTTCACTTGCCATTGCAGA---AGGTCCTGTTAT 255

Qy   74 PheGlyGluHisLeuGlyGlyThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
          |||:||||||||||||::::||
Db   256 TCAACAGAATTATTCCTATGGGACCTGCACTGTCATGGGTTATTAACCACACAGAATCTGCTGC 315

RESULT 5
US-08-158-189-4
; Sequence 4, Application US/0815189
; Patent No. 5641497
GENERAL INFORMATION:
  APPLICANT: Bevins, Charles L.
  APPLICANT: Jones, Douglas E.
  TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
  TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
  NUMBER OF SEQUENCES: 51
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
    STREET: One Liberty Place - 46th Floor
    CITY: Philadelphia
    STATE: PA
    COUNTRY: U.S.A.
    ZIP: 19103
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/158,189
    FILING DATE:
  PRIOR APPLICATION DATA:
    CLASSIFICATION: 435
    APPLICATION NUMBER: US/07/888,232
    FILING DATE:
  ATTORNEY/AGENT INFORMATION:
    NAME: Johnson, Philip S.
    REGISTRATION NUMBER: 27,200
    REFERENCE/DOCKET NUMBER: CH-0219
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 215-568-3100
    TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 4:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 424 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: unknown
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 10..294
; US-08-158-189-4

Alignment Scores:

Pred. No.:   1.12e-17   Length: 424
Score: 207.00   Matches: 44
Percent Similarity: 59.14%   Conservative: 11
Best Local Similarity: 47.31%   Mismatches: 38
Query Match: 40.27%   Indels: 0
DB: 1   Gaps: 0

US-10-045-180A-3 (1-94) x US-08-158-189-4 (1-424)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
          |||||||||:|:::   :::||:|||||||||||||||||:|||||
Db   10 ATGAGGACCATCGCCATCCTTTGCTGCCATTCTCCTTGGTGGTGCCCTGCAGGCCTGGGCCGAG 69

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
          :::||:||||| ||||||||:::
Db   70 TCACTCCAGGAAAGAGGCTCATGAGGGTCAGCCACCCGCAGGAGTCTGGGGAAGACAACCAG 129

Qy   41 AspValValIleTyrPheSerGlyAlaAspAspAspProSerGlnLeuGlnValProGlySerThr 60
          |||:| |||:||||| ::|||||:::
Db   130 GACCTTGCTATCTCCTTTGCAGGGAAAATGGACACTCTCGTCTTAGAACCTGAGTTCCTGAG 189

Qy   61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly 80
          |||: ||||| :::|||||| :::|||
Db   190 GCAAGAGCCAACCTGCTATTGGCCGACCGCCGGTTGTGCTACCCGTGAGTCCCTCCGGG 249

Qy   81 ThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
          |||:|||| :::||||
Db   250 GTGTGTGAAATCAGTGGCGCCCCTCACAGACTCTGCTGT 288

RESULT 6
US-08-486-013-46
; Sequence 46, Application US/0848601
; Patent No. 5731149
GENERAL INFORMATION:
  APPLICANT: Selsted, Michael E.
  APPLICANT: Ouellette, Andre J.
  TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
  TITLE OF INVENTION: of Their Use
  NUMBER OF SEQUENCES: 70
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Campbell and Flores
    STREET: 4370 La Jolla Village Drive, Suite 700
    CITY: San Diego
    STATE: California
    COUNTRY: USA
    ZIP: 92122
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/486,013
    FILING DATE:
  PRIOR APPLICATION DATA:
    CLASSIFICATION: 530
    APPLICATION NUMBER: US 08/342,268
    FILING DATE: 18-NOV-1994
    APPLICATION NUMBER: US 07/930,649
    FILING DATE: 14-AUG-1992
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/889,020
    FILING DATE: 26-MAY-1992
  ATTORNEY/AGENT INFORMATION:
    NAME: Campbell, Cathryn A.
    REGISTRATION NUMBER: 31,815
    REFERENCE/DOCKET NUMBER: P-UC 1206
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (619) 535-9001
    TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 46:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 420 base pairs

Left column (RESULT 7)

```
;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-486-013-46

Alignment Scores:
Pred. No.:                 2.03e-13
Score:                     173.50       Length:        420
Percent Similarity:        55.91%       Matches:        38
Best Local Similarity:     40.86%       Conservative:   14
Query Match:               33.75%       Mismatches:     40
DB:                                     Indels:          1
                                        Gaps:            1

US-10-045-180A-3 (1-94) x US-08-486-013-46 (1-420)

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
          |||::|:||||
Db   44 ATGAAGACACTAATCCTCCTCCTGTCCTTGGGTCCTTCAGGTCCAGGCTGAT 103

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnHisGlnProProAlaAspArgGln  40
          |||       ::|||
Db  104 CCTATCCAAAACACAGATGAAGAGACTAAAACTGAGGAGCAGCCACGGGGAAGACGACCAG 163

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr  60
          ::|
Db  164 GCCGTATCGTGTCTCCCTTTGGAGACCCAGAAGGCTCTTCTCTTCTTCAA--GAGAAATCGGTTG 220

Qy   61 LysGlyIleLeuIleCysHisCysSerArgValLeuTyrCysIlePheGlyGluHisGlyGly  80
          ::|
Db  221 AGAGATCTGGTATGCTATTGTAGACAAGAGGACTGCAAAAGAAGAGAACACATGAATGGG 280

Qy   81 ThrCysPheLeuLeuGlyGlyArgTyrProIleCysCys  93
          |||||          |||
Db  281 ACCTGCAGAAAGGGTCATTTAATGTACACGGCTCTGCTGT 319
```

```
RESULT 7
US-08-482-279-46
Sequence 46, Application US/0848279
Patent No. 5840498
GENERAL INFORMATION:
APPLICANT:  Selsted, Michael E.
APPLICANT:  Ouellette, Andre J.
TITLE OF INVENTION:  Antibiotic Cryptdin Peptides and Methods
TITLE OF INVENTION:  of Their Use
NUMBER OF SEQUENCES:  70
CORRESPONDENCE ADDRESS:
   ADDRESSEE:  Campbell and Flores
   STREET:  4370 La Jolla Village Drive, Suite 700
   CITY:  San Diego
   STATE:  California
   COUNTRY:  USA
   ZIP:  92122
COMPUTER READABLE FORM:
   MEDIUM TYPE:  Floppy disk
   COMPUTER:  IBM PC compatible
   OPERATING SYSTEM:  PC-DOS/MS-DOS
   SOFTWARE:  Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER:  US/08/482,279
   FILING DATE:
PRIOR APPLICATION DATA:
   CLASSIFICATION:  530
PRIOR APPLICATION DATA:
   APPLICATION NUMBER:  US 08/342,268
   FILING DATE:  18-NOV-1994
   APPLICATION NUMBER:  US 07/930,649
   FILING DATE:  14-AUG-1992
PRIOR APPLICATION DATA:
   APPLICATION NUMBER:  US 07/889,020
   FILING DATE:  26-MAY-1992
ATTORNEY/AGENT INFORMATION:
   NAME:  Campbell, Cathryn A.
   REGISTRATION NUMBER:  31,815
   REFERENCE/DOCKET NUMBER:  P-UC 1206
```

Right column (RESULT 8)

```
;
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (619) 535-9001
;   TELEFAX:  (619) 535-8949
; INFORMATION FOR SEQ ID NO: 46:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 420 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
; US-08-482-279-46

Alignment Scores:
Pred. No.:                 2.03e-13
Score:                     173.50       Length:        420
Percent Similarity:        55.91%       Matches:        38
Best Local Similarity:     40.86%       Conservative:   14
Query Match:               33.75%       Mismatches:     40
DB:                                     Indels:          1
                                        Gaps:            1

US-10-045-180A-3 (1-94) x US-08-482-279-46 (1-420)

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
          |||::|:||||
Db   44 ATGAAGACACTAATCCTCCTCCTGTCCTTGGGTCCTTCAGGTCCAGGCTGAT 103

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnHisGlnProProAlaAspArgGln  40
          |||
Db  104 CCTATCCAAAACACAGATGAAGAGACTAAAACTGAGGAGCAGCCACGGGGAAGACGACCAG 163

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr  60
          :::
Db  164 GCCGTATCGTGTCTCCCTTTGGAGACCCAGAAGGCTCTTCTCTTCTTCAA--GAGAAATCGGTTG 220

Qy   61 LysGlyIleLeuIleCysHisCysSerArgValLeuTyrCysIlePheGlyGluHisGlyGly  80
          ||||||||:::
Db  221 AGAGATCTGGTATGCTATTGTAGACAAGAGGACTGCAAAAGAAGAGAACACATGAATGGG 280

Qy   81 ThrCysPheLeuLeuGlyGlyArgTyrProIleCysCys  93
          :::           ::::
Db  281 ACCTGCAGAAAGGGTCATTTAATGTACACGGCTCTGCTGT 319
```

```
RESULT 8
US-08-342-268-46
Sequence 46, Application US/08342268
Patent No. 5844072
GENERAL INFORMATION:
APPLICANT:  Selsted, Michael E.
APPLICANT:  Ouellette, Andre J.
TITLE OF INVENTION:  Antibiotic Cryptdin Peptides and Methods
TITLE OF INVENTION:  of Their Use
NUMBER OF SEQUENCES:  70
CORRESPONDENCE ADDRESS:
   ADDRESSEE:  Campbell and Flores
   STREET:  4370 La Jolla Village Drive, Suite 700
   CITY:  San Diego
   STATE:  California
   COUNTRY:  USA
   ZIP:  92122
COMPUTER READABLE FORM:
   MEDIUM TYPE:  Floppy disk
   COMPUTER:  IBM PC compatible
   OPERATING SYSTEM:  PC-DOS/MS-DOS
   SOFTWARE:  Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER:  US/08/342,268
   FILING DATE:  18-NOV-1994
   CLASSIFICATION:  435
PRIOR APPLICATION DATA:
   APPLICATION NUMBER:  US 07/930,649
   FILING DATE:  14-AUG-1992
PRIOR APPLICATION DATA:
   APPLICATION NUMBER:  US 07/889,020
   FILING DATE:  26-MAY-1992
```

~ ATTORNEY/AGENT INFORMATION:
~ NAME: Campbell, Cathryn A.
~ REGISTRATION NUMBER: 31,815
~ REFERENCE/DOCKET NUMBER: P-UC 1206
~ TELECOMMUNICATION INFORMATION:
~ TELEPHONE: (619) 535-9001
~ TELEFAX: (619) 535-8949
~ INFORMATION FOR SEQ ID NO: 46:
~ SEQUENCE CHARACTERISTICS:
~ LENGTH: 420 base pairs
~ TYPE: nucleic acid
~ STRANDEDNESS: double
~ TOPOLOGY: linear
~ US-08-342-268-46

US-10-045-180A-3 (1-94) x US-08-342-268-46 (1-420)

Alignment Scores:

Pred. No.:	2.03e-13	Length:	420
Score:	173.50	Matches:	38
Percent Similarity:	55.91%	Conservative:	14
Best Local Similarity:	40.86%	Mismatches:	40
Query Match:	33.75%	Indels:	1
DB:	2	Gaps:	1

```
Qy     1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
           ||:||||||:|||||||              |||||||||:||
Db    44 ATGAAGACACTAATCCTCCTCTCTGCCCTCGTCCTGCCTTCCAGGTCGTCCAGGCTGAT 103

Qy    21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProAlaLeuAspAspGln  40
           |||:|||||                     |||||||||||||||||
Db   104 CCTATCCAAAACACAGATGAAGAGACTAAAACTGAGGAGGCAGCCGGGGAAGACGACCAG 163

Qy    41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr  60
           ::|::
Db   164 GCCGTATCTGTCTCCTTTGGAGACCCAGAAGGCTTCTCTTCAA---GAGGAATCGTTG 220

Qy    61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly  80
           |||:||:||||||||             ||:::                 |||
Db   221 AGAGATCTCGGTATGCTATTGTAGAACAAGAGGCTGCAAAAGAGAGAGAGACACATGAAATGGG 280

Qy    81 ThrCysPheIleLeuGlyGluArgTyrProIleCysCys  93
           ||||||             :::||||||
Db   281 ACCTGCAGAAAGGGTCATTTAAATGTACACGCTCTGCTGT 319
```

~ US-10-045-180A-3
~ US-08-342-268-46

RESULT 9
US-09-015-968-46
; Sequence 46, Application US/09015968
; Patent No. 6057425
GENERAL INFORMATION:
~ APPLICANT: Selsted, Michael E.
~ APPLICANT: Ouellette, Andre J.
~ TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
~ TITLE OF INVENTION: of Their Use
~ NUMBER OF SEQUENCES: 70
~ CORRESPONDENCE ADDRESS:
~ ADDRESSEE: Campbell and Flores
~ STREET: 4370 La Jolla Village Drive, Suite 700
~ CITY: San Diego
~ STATE: California
~ COUNTRY: USA
~ ZIP: 92122
~ COMPUTER READABLE FORM:
~ MEDIUM TYPE: Floppy disk
~ COMPUTER: IBM PC compatible
~ OPERATING SYSTEM: PC-DOS/MS-DOS
~ SOFTWARE: Patentin Release #1.0, Version #1.25
~ CURRENT APPLICATION DATA:
~ APPLICATION NUMBER: US/09/015,968
~ FILING DATE:
~ CLASSIFICATION:
~ PRIOR APPLICATION DATA:
~ APPLICATION NUMBER: US 08/482,279

~ FILING DATE: 07-JUN-1995
~ PRIOR APPLICATION DATA:
~ APPLICATION NUMBER: US 08/342,268
~ FILING DATE: 18-NOV-1994
~ PRIOR APPLICATION DATA:
~ APPLICATION NUMBER: US 07/930,649
~ FILING DATE: 14-AUG-1992
~ PRIOR APPLICATION DATA:
~ APPLICATION NUMBER: US 07/889,020
~ FILING DATE: 26-MAY-1992
~ ATTORNEY/AGENT INFORMATION:
~ NAME: Campbell, Cathryn A.
~ REGISTRATION NUMBER: 31,815
~ REFERENCE/DOCKET NUMBER: P-UC 3003
~ TELECOMMUNICATION INFORMATION:
~ TELEPHONE: (619) 535-9001
~ TELEFAX: (619) 535-8949
~ INFORMATION FOR SEQ ID NO: 46:
~ SEQUENCE CHARACTERISTICS:
~ LENGTH: 420 base pairs
~ TYPE: nucleic acid
~ STRANDEDNESS: double
~ TOPOLOGY: linear
~ US-09-015-968-46

US-10-045-180A-3 (1-94) x US-09-015-968-46 (1-420)

Alignment Scores:

Pred. No.:	2.03e-13	Length:	420
Score:	173.50	Matches:	38
Percent Similarity:	55.91%	Conservative:	14
Best Local Similarity:	40.86%	Mismatches:	40
Query Match:	33.75%	Indels:	1
DB:	3	Gaps:	1

```
Qy     1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuLeuValAlaLeuGlnAlaTrpAlaGlu  20
           ||:||||||:|||||||              |||||||||:||
Db    44 ATGAAGACACTAATCCTCCTCTCTGCCCTCGTCCTGCCTTCCAGGTCGTCCAGGCTGAT 103

Qy    21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProAlaLeuAspAspGln  40
           |||:|||||                     |||||||||||||||||
Db   104 CCTATCCAAAACACAGATGAAGAGACTAAAACTGAGGAGGCAGCCGGGGAAGACGACCAG 163

Qy    41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr  60
           ::|::
Db   164 GCCGTATCTGTCTCCTTTGGAGACCCAGAAGGCTTCTCTTCAA---GAGGAATCGTTG 220

Qy    61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly  80
           |||:||:||||||||             ||:::                 |||
Db   221 AGAGATCTCGGTATGCTATTGTAGAACAAGAGGCTGCAAAAGAGAGAGAGACACATGAAATGGG 280

Qy    81 ThrCysPheIleLeuGlyGluArgTyrProIleCysCys  93
           ||||||             :::||||||
Db   281 ACCTGCAGAAAGGGTCATTTAAATGTACACGCTCTGCTGT 319
```

~ US-10-045-180A-3
~ US-09-015-968-46

RESULT 10
US-09-397-386-46
; Sequence 46, Application US/09397386
; Patent No. 6300470
GENERAL INFORMATION:
~ APPLICANT: Selsted, Michael E.
~ APPLICANT: Ouellette, Andre J.
~ TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
~ TITLE OF INVENTION: of Their Use
~ NUMBER OF SEQUENCES: 70
~ CORRESPONDENCE ADDRESS:
~ ADDRESSEE: Campbell and Flores
~ STREET: 4370 La Jolla Village Drive, Suite 700
~ CITY: San Diego
~ STATE: California
~ COUNTRY: USA
~ ZIP: 92122
~ COMPUTER READABLE FORM:

Column 1

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/015,968
FILING DATE:
APPLICATION NUMBER: US 08/482,279
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/342,268
FILING DATE: 18-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,649
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,020
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UC 3003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 420 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-09-397-386-46

Alignment Scores:
Pred. No.:
Score: 2.03e-13   Length: 420
Percent Similarity: 55.91%   Matches: 38
Best Local Similarity: 40.86%   Conservative: 14
Query Match: 33.75%   Mismatches: 40
                                          Indels: 1
                                          Gaps: 1

US-10-045-180a-3 (1-94) x US-09-397-386-46 (1-420)

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaLeuTrpAlaGlu 20
        |||||||||||||||||||| |||||||||||||||||||||::::
Db   44 ATGAAGACAACTAATCCTCTCTTCCTGCCCGTTCGTTCCAGTCCAGGACCAG 103

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
        ||||||||||||
Db  104 CCTATCCAAAACACAGATGAAGAGCAGCCAGGCCAGGGAGAAGACGAACCAG 163

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
        |||    |||
Db  164 GCCGTTATCTATCTCCTTTGGAGACCCAGAAGGCTTCTCTTCAA---GAGGAATCGTTG 220

Qy   61 LysGlyLeuIleCysHisCysArgValLeuPheGlyHisIleLeuGlyGly 80
        |||::: |||
Db  221 AGAGATCTGGTATCGTATTGTAGAACAAGAGGCTGCAAAAGAAGAGAACACATGAAATGGG 280

Qy   81 ThrCysPheLeuLeuGlyGluArgTyrProIleCysCys 93
        |||||
Db  281 ACCTGCAGAAAGGTCATTTAATCTACACGCTCTGCTGT 319
```

RESULT 11
US-08-013-54
Sequence 54, Application US/08486013
Patent No. 5731149
GENERAL INFORMATION:
APPLICANT: Selsted, Michael E.

Column 2

APPLICANT: Ouellette, Andre J.
TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
TITLE OF INVENTION: of Their Use
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,013
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/342,268
FILING DATE: 18-NOV-1994
APPLICATION NUMBER: US 07/930,649
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,020
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UC 1206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-08-486-013-54

Alignment Scores:
Pred. No.:
Score: 3.43e-13   Length: 401
Percent Similarity: 55.91%   Matches: 38
Best Local Similarity: 40.86%   Conservative: 14
Query Match: 33.37%   Mismatches: 40
                                          Indels: 1
                                          Gaps: 1

US-10-045-180a-3 (1-94) x US-08-486-013-54 (1-401)

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaLeuTrpAlaGlu 20
        |||||||||||||||||||| |||||||||||||||||::::
Db   26 ATGAAGACAACTAGTCCTCTCTTCCTCGCCCGTTGTCTGGCCGTTCCAGGTCCAGGGCTGAT 85

Qy   21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
        ||||||||||||
Db   86 CCTATCCAAAACACAGATGAAGAGACTAAAAACTGAGGAGCAGCCAGGGAGGAGACGAACCAG 145

Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
        |||    :::
Db  146 GCCGTTATCTATCTCTCTTTGGAGACCCCAGAAGGCGCCCTTCTCTTCAA---GAGGAATCGTTG 202

Qy   61 LysGlyLeuIleCysHisCysArgValLeuPheGlyHisIleLeuGlyGly 80
        |||::: |||
Db  203 AGAGATCTGGTATCGTATTGTAGAAAAGAGGCTGCAAAAGAAGAGAACACATAAATGGG 262

Qy   81 ThrCysPheLeuLeuGlyGluArgTyrProIleCysCys 93
        |||||
Db  263 ACCTGCAGAAAGGTCATTTATTGTACACTCTCTGCTGT 301
```

RESULT 12
US-08-482-279-54

; Sequence 54, Application US/08482279
; Patent No. 5840498
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
; TITLE OF INVENTION: of Their Use
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,279
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/342,268
; FILING DATE: 18-NOV-1994
; APPLICATION NUMBER: US 07/930,649
; FILING DATE: 14-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/889,020
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UC 1206
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

US-08-482-279-54

Alignment Scores:
Pred. No.:
Score: 3.43e-13
Percent Similarity: 171.50
Percent Similarity: 55.91%
Best Local Similarity: 40.86%
Query Match: 33.37%
DB:

```
                                                Length:        401
                                                Matches:       38
                                                Conservative:  14
                                                Mismatches:    40
                                                Indels:        1
                                                Gaps:          1
```

US-10-045-180A-3 (1-94) x US-08-482-279-54 (1-401)

```
Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaThrProAlaGlu 20
         ||||||  ||||||       ||||||| ||:|||    ||||   |:||
Db  26 ATGAAGACAGACTAGTCCTCCTCTGCCCTTGTCCCTTTCCAGGTCCAGGCTGAT 85

Qy  21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
         ||||||    ||:|||                     |||||| ||:|:
Db  86 CCTATCCAAAACACAGATGAAGAGACTAAAACTGAGGAGGAAGACGACCAG 145

Qy  41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
         |||   |:|||       ||||||||||||
Db 146 GCCGTACTGTGTCCTTTGGACCCAGAGGGCTCTTCTTCAA---GAGGAATCGTTG 202

Qy  61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyHisLeuGlyGly 80
         |||  |||::|||:||||||:||||
```

Db 203 AGAGATCTGGTATGCTATTGTATTAGAAAAAGAGGCTGCAAAAGAAGAGAACACATAAATGGG 262

Qy 81 ThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
         |||||| :::|||||||
Db 263 ACCTGCAGAAAAGGTCATTATTGTGTACACTCCGTCGT 301

RESULT 13
US-08-342-268-54

; Sequence 54, Application US/08342268
; Patent No. 5844072
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
; TITLE OF INVENTION: of Their Use
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/342,268
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,649
; FILING DATE: 14-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/889,020
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UC 1206
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

US-08-342-268-54

Alignment Scores:
Pred. No.:
Score: 3.43e-13
Percent Similarity: 171.50
Percent Similarity: 55.91%
Best Local Similarity: 40.86%
Query Match: 33.37%
DB:

```
                                                Length:        401
                                                Matches:       38
                                                Conservative:  14
                                                Mismatches:    40
                                                Indels:        1
                                                Gaps:          1
```

US-10-045-180A-3 (1-94) x US-08-342-268-54 (1-401)

```
Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaThrProAlaGlu 20
         ||||||  ||||||       ||||||| ||:|||    ||||   |:||
Db  26 ATGAAGACAGACTAGTCCTCCTCTGCCCTTGTCCCTTTCCAGGTCCAGGCTGAT 85

Qy  21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
         ||||||    ||:|||                     |||||| ||:|:
Db  86 CCTATCCAAAACACAGATGAAGAGACTAAAACTGAGGAGGAAGACGACCAG 145

Qy  21 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
Db  26

Qy  41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
         |||   |:|||       ||||||||||||
```

RESULT 14
US-09-015-968-54
; Sequence 54, Application US/09015968
; Patent No. 6057425
; GENERAL INFORMATION:
; APPLICANT: Ouellette, Andre J.
; APPLICANT: Selsted, Michael E.
; TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
; TITLE OF INVENTION: of Their Use
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/015,968
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/889,020
; FILING DATE: 26-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,279
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/342,268
; FILING DATE: 18-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,649
; FILING DATE: 14-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/015,968
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UC 3003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

US-09-015-968-54

Alignment Scores:
Pred. No.:
Score: 3.43e-13
Score: 171.50   Length: 401
Percent Similarity: 55.91%   Matches: 38
Best Local Similarity: 40.86%   Conservative: 14
Query Match: 33.37%   Mismatches: 40
DB: Indels: 1
                                                Gaps: 1

US-10-045-180A-3 (1-94) x US-09-015-968-54 (1-401)

Db   146 GCCGTATCTGTCTCCTTTGGAGACCCAGAAGGCTCTTCTTCTTCAA---GAGGAATCGTTG 202
Qy   61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyHisLeuGlyGly 80
Db   203 AGAGATCTGGTATGCTATTGTAGAAAAGAGGCTGCAAAAAGAAGAGAACACATAAAATGG 262
Qy   81 ThrCysPheIleLeuGlyGlyArgTyrProIleCysCys 93
Db   263 ACCTGCAGAAAGGGTCATTTATTGTACACTCTCTGCTGT 301

RESULT 15
US-09-397-386-54
; Sequence 54, Application US/09397386
; Patent No. 6300470
; GENERAL INFORMATION:
; APPLICANT: Ouellette, Andre J.
; APPLICANT: Selsted, Michael E.
; TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
; TITLE OF INVENTION: of Their Use
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/397,386
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/015,968
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,279
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/342,268
; FILING DATE: 18-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,649
; FILING DATE: 14-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/889,020
; FILING DATE: 26-MAY-1992
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UC 3003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaArgProAlaGlu 20
Db   26 ATGAAGACACTAGTCCTGCTTTCTGCCTTTCTGCTGGTCGCTCTGCAGGCTCGAT 85
Qy   21 ProLeuGlnAlaArgProAlaHisGlnMetProAlaGlnInLysGlnProProAlaSpAspGln 40
Db   86 CCTATCCAAAACACAGATGAAGAGACTAAAACTGAGAGAGCAGCCGGGGAAGACGACAG 145
Qy   41 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
Db   146 GCCGTATCTGTCTCCTTTGGAGACCCAGAAGGCTCTTCTTCTTCAA---GAGGAATCGTTG 202
Qy   61 LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyHisLeuGlyGly 80
Db   203 AGAGATCTGGTATGCTATTGTAGAAAAGAGGCTGCAAAAAGAAGAGAACACATAAAATGG 262
Qy   81 ThrCysPheIleLeuGlyGlyArgTyrProIleCysCys 93
Db   263 ACCTGCAGAAAGGGTCATTTATTGTACACTCTCTGCTGT 301

US-09-397-386-54

Alignment Scores:
Pred. No.:
Score: 3.43e-13   Length: 401
Percent Similarity: 171.50   Matches: 38
Best Local Similarity: 55.91%   Conservative: 14
Query Match: 40.86%   Mismatches: 40
DB: 33.37%   Indels: 1
                                4   Gaps: 1

US-10-045-180A-3 (1-94) x US-09-397-386-54 (1-401)

Qy   1   MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaGlu 20
           |||:|||||||||||||||||||||||||||||||||:::
Db   26   ATGAAGACACTAGTCCTCCTTCTCTGCCCCTTGTCCTGGCCTTTCCAGGTCCAGGCTGAT 85

Qy   21   ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 40
           |||::|||:||| :::::||||||||||
Db   86   CCTATCCAAAAACACAGAGAAGAGACTAAAACTGAGGAGCAGCCCAGGGGGAAGAGACCAG 145

Qy   41   AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 60
           |||:   :::|||
Db   146 GCCGTATCTGTCTCTCCTTTGGAGACCCAGAAGGGCTTCTTCTTCTTCAA---GAGGAATCGTTG 202

Qy   61   LysGlyLeuIleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGly 80
           :::|||||||||:|||::||| |||||::::
Db   203 AGAGATCTGGTATCGTATTGTAGAAAAAGAGGGCTGCAAAAGACAGAGAACACATAAAATGGG 262

Qy   81   ThrCysPheIleLeuGlyGluArgTyrProIleCysCys 93
           |||||| :::||||||||
Db   263 ACCTGCAGAAAGGGTCATTTATTGTACACTCTCTGCTGT 301

Search completed: December 17, 2003, 14:53:59
Job time : 97 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 14:50:33 ; Search time 543.622 Seconds
(without alignments)
1429.821 Million cell updates/sec

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Title:
Perfect score: 90
Sequence: 1 MRTLTLLSAFLLVALQAWA 19
US-10-045-180A-4

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Searched: 2888711 seqs, 2045813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145023_12528/app_query.fasta_1.860
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10045180@CGN_1_1_5283 @runat_17122003_145023_12528 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
```
 1: gb_ba:*
 2: gb_bl:*
 3: gb_in:*
 4: gb_om:*
 5: gb_ov:*
 6: gb_pat:*
 7: gb_ph:*
 8: gb_pl:*
 9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_ba:*
16: gb_fun:*
17: em_hum:*
18: em_in:*
19: em_mm:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Pred.	Score	% Query Match	Length	DB	ID	Description
1		90	100.0	453	6	A98571	A98571 Sequence 2
2		90	100.0	453	6	BD074746	BD074746 Human def
3		90	100.0	4415	6	A98570	A98570 Sequence 1
4		90	100.0	4415	6	BD074745	BD074745 Human def
5	c	85	94.4	1285444	2	AF238378	AF238378 Homo sapi
6	c	65	72.2	232180	2	AC021883	AC021883 Homo sapi
7		64	71.1	398	2	AX405993	AX405993 Sequence
8		64	71.1	452	6	AR270635	AR270635 Sequence
9		64	71.1	452	6	I49588	I49588 Sequence 6
10		64	71.1	452	6	HUMDFSNSIX	M98331 Homo sapien
11		64	71.1	3060	6	HSU33317	U33317 Human defen
12		64	71.1	79730	9	AF314060	AF314060 Homo sapi
13	c	64	71.1	150940	9	AF233439	AF233439 Homo sapi
14	c	64	71.1	158653	9	AF200455	AF200455 Homo sapi
15	c	64	71.1	168702	9	AC079018	AC079018 Homo sapi
16	c	64	71.1	169044	9	AF205406	AF205406 Homo sapi
17		63	70.0	195	2	I49590	I49590 Sequence 9
18		63	70.0	202	6	I49589	I49589 Sequence 8
19		63	70.0	479	4	RABNFCS4A	L10841 Oryctolagus
20		63	70.0	495	4	RABMCP1A	M28880 Rabbit macr
21		63	70.0	500	4	S55582	S55582 Oryctolagus
22		63	70.0	532	4	RABMCP2A	M28884 Rabbit macr
23		63	70.0	2097	4	RABMCP1AA	M28072 Rabbit macr
24		63	70.0	2133	4	RABMCP1AB	M28073 Rabbit macr
25		61	67.8	110	6	AR021138	AR021138 Sequence
26		61	67.8	110	6	I33694	I33694 Sequence 9
27		61	67.8	110	6	I74645	I74645 Sequence 9
28		61	67.8	110	6	I79691	I79691 Sequence 9
29		61	67.8	203	6	I49591	I49591 Sequence 10
30		61	67.8	203	6	I49592	I49592 Sequence 11
31		61	67.8	218	6	I49596	I49596 Sequence 38
32		61	67.8	347	6	AF188268	AF188268 Macaca mu
33		61	67.8	347	6	AF188269	AF188269 Macaca mu
34		61	67.8	347	6	AF188270	AF188270 Macaca mu
35		61	67.8	433	9	AF184159	AF184159 Macaca mu
36		61	67.8	435	9	AF184160	AF184160 Macaca mu
37		61	67.8	448	9	HUMDEFP1A	M21130 Human neutr
38		61	67.8	451	9	HUMDEFB	M23281 Human defen
39		61	67.8	464	9	HSDEF3	X13621 Human HNP-3
40		61	67.8	478	9	BD028213	BD028213 Sequence
41		61	67.8	482	6	HSHP1	X52053 Human mRNA
42		61	67.8	498	6	BD174049	BD174049 Method of
43		61	67.8	498	6	HUMDEF1AA	M26602 Human defen
44		61	67.8	514	2	AX405718	AX405718 Sequence
45		61	67.8	556	9	BC027917	BC027917 Homo sapi

RESULT 1

ALIGNMENTS

RESULT 1

A98571
```
LOCUS       A98571        453 bp    DNA      linear    PAT 26-JAN-2000
DEFINITION  Sequence 2 from Patent WO9911663.
ACCESSION   A98571
VERSION     A98571.1  GI:6781627
KEYWORDS
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Bougueleret,L. and Chumakov,I.
  TITLE     HUMAN DEFENSIN DEF-X, GENE AND DNAc, COMPOSITION CONTAINING SAME
            AND DIAGNOSTIC AND THERAPEUTIC APPLICATIONS
  JOURNAL   Patent: WO 9911663-A 11-MAR-1999;
            BOUGUELERET LYDIE (FR); CHUMAKOV ILYA (FR)
FEATURES             Location/Qualifiers
     source          1..453
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT    107 a     121 c      97 g     128 t
ORIGIN
```

US-10-045-180A-4 (1-19) x A98571 (1-453)

```
Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52 ATGAGGACCCTCACCCTCCTCTCTGCCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCA 108
```

Alignment Scores:
```
Pred. No.:
Score:                  3.61e-07    Length:         453
Percent Similarity:     90.00       Matches:         19
Best Local Similarity: 100.00%      Conservative:     0
Query Match:           100.00%      Mismatches:       0
DB:                     6           Indels:           0
                                    Gaps:             0
```

RESULT 2
BD074746
```
LOCUS       BD074746      453 bp    DNA      linear    PAT 27-AUG-2002
DEFINITION  Human defensin polypeptide Def-X, genome DNA and cDNA, composition
            containing the same, and application to diagnosis and remedy
            treatment.
ACCESSION   BD074746
VERSION     BD074746.1  GI:22620349
KEYWORDS    JP 2001514264-A/2.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Bougueleret,L. and Shmacov,I.
  TITLE     Human defensin polypeptide Def-X, genome DNA and cDNA, composition
            containing the same, and application to diagnosis and remedy
  JOURNAL   Patent: JP 2001514264-A 2 11-SEP-2001;
            GENSET
OS   Homo sapiens (human)
PN   JP 2001514264-A/2
PD   11-SEP-2001
PF   28-AUG-1998 JP 2000508701
PR   29-AUG-1997 FR 97/10823
PI   LYDIE BOUGUELERET, ILYA SHMACOV
PI   LYDIR BOUGUELERET, ILYA SHMACOV
PC   C07K14/435,A01N43/50,A61K7/00,A61K38/00,A61P29/00,
PC   A61P35/00,
PC   A61P37/02,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N15/09, PC
PC   C12P21/02,
PC   C12P21/08,C12Q1/68,G01N33/53,A61K37/02,A61K37/02,C12N15/00 PC
PC   C12P21/08,C12Q1/68,G01N33/53,A61K37/02,C12N15/00 CC
COMMENT     Strandedness: Double;
CC          Topology: Linear;
CC          Human defensin polypeptide Def-X, genome DNA and cDNA, CC
CC          composition and application to diagnosis and remedy CC
CC          containing the same, and application to diagnosis and remedy
CC          treatment
```

```
FEATURES             Location/Qualifiers
FH   Key             Location/Qualifiers
FT   source          1..453
FT                   /organism="Homo sapiens (human)".
     source          1..453
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT    107 a     121 c      97 g     128 t
ORIGIN
```

US-10-045-180A-4 (1-19) x BD074746 (1-453)

```
Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52 ATGAGGACCCTCACCCTCCTCTCTGCCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCA 108
```

Alignment Scores:
```
Score:                  3.61e-07    Length:         453
Percent Similarity:     90.00       Matches:         19
Best Local Similarity: 100.00%      Conservative:     0
Query Match:           100.00%      Mismatches:       0
DB:                     6           Indels:           0
                                    Gaps:             0
```

RESULT 3
A98570
```
LOCUS       A98570        453 bp    DNA      linear    PAT 26-JAN-2000
DEFINITION  Sequence 1 from Patent WO9911663.
ACCESSION   A98570
VERSION     A98570.1  GI:6781626
KEYWORDS
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4415)
  AUTHORS   Bougueleret,L. and Chumakov,I.
  TITLE     HUMAN DEFENSIN DEF-X, GENE AND DNAc, COMPOSITION CONTAINING SAME
            AND DIAGNOSTIC AND THERAPEUTIC APPLICATIONS
  JOURNAL   Patent: WO 9911663-A 1 11-MAR-1999;
            BOUGUELERET LYDIE (FR); CHUMAKOV ILYA (FR)
FEATURES             Location/Qualifiers
     source          1..4415
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
     exon            1836..1874
     exon            3394..3577
     exon            4161..4380
     polyA_site      4374..4379
BASE COUNT   1128 a    1109 c     912 g    1252 t     14 others
ORIGIN
```

US-10-045-180A-4 (1-19) x A98570 (1-4415)

```
Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52 ATGAGGACCCTCACCCTCCTCTCTGCCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCA 108
```

Alignment Scores:
```
Pred. No.:
Score:                  3.93e-06    Length:        4415
Percent Similarity:     90.00       Matches:         19
Best Local Similarity: 100.00%      Conservative:     0
Query Match:           100.00%      Mismatches:       0
DB:                     6           Indels:           0
                                    Gaps:             0
```

RESULT 4
BD074745
```
LOCUS       BD074745      4415 bp   DNA      linear    PAT 27-AUG-2002
DEFINITION  Human defensin polypeptide Def-X, genome DNA and cDNA, composition
            containing the same, and application to diagnosis and remedy
            treatment.
```

```
ACCESSION   BD074745
VERSION     BD074745.1  GI:22620348
KEYWORDS    JP 2001514264-A/1.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4415)
AUTHORS     Bougueleret,L. and Shmacov,I.
TITLE       Human defensin polypeptide Def-X, genome DNA and cDNA, composition
            containing the same, and application to diagnosis and remedy
JOURNAL     Patent: JP 2001514264-A 1 11-SEP-2001;
            GENSET
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001514264-A/1
            PD   11-SEP-2001
            PF   28-AUG-1998 JP 2000508701
            PR   29-AUG-1997 FR 97/10823
            PI   LYDIE BOUGUELERET, ILYA SHMACOV
            PC   C07K14/435,A01N43/50,A01N63/00,A61K7/00,A61K38/00,A61P29/00,
            PC   A61P35/00,
            PC   A61P37/02,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N15/09, PC
            PC   C12P21/08,C12Q1/68,G01N33/53,A61K37/02,C12N15/00 CC
            PC   C12P21/02,
            Strandedness: Double;
            Topology: Linear;
            Human defensin polypeptide Def-X, genome DNA and cDNA, CC
            composition
            containing the same, and application to diagnosis and remedy CC
            treatment
FEATURES             Location/Qualifiers
     source          1..4415
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     FH
     CC
     CC
     FT   Key             Location/Qualifiers
     FT   exon            1836..1874
     FT   exon            3394..3577
     FT   exon            4161..4380
     FT   CDS             3406..3408
     FT   CDS             4276..4278.
BASE COUNT   1128 a  1109 c  912 g  1252 t  14 others
ORIGIN

US-10-045-180A-4 (1-19) x BD074745 (1-4415)

RESULT 5

Alignment Scores:
Pred. No.:            1
Score:               3.93e-06
Percent Similarity:  90.00
Best Local Similarity: 100.00%
Query Match:         100.00%

Length:              4415
Matches:             19
Conservative:        0
Mismatches:          0
Indels:              0
Gaps:                0

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuValAlaLeuGlnAlaTrpAla 19
Db 3406 ATGAGGACCCTCACCCTCCTCTCGCTTTCTCCCTGGTGGCCCTTCAGGCCTGGGCA 3462
```

```
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 125544)
AUTHORS     Wen,G., Schutte,B., Malik,M., Peng,J.Hong., McCray,P. and
            Rosenthal,A.
TITLE       Chromosome 8 genomic sequence
JOURNAL     Unpublished
REFERENCE   3  (bases 1 to 125544)
AUTHORS     Schudy,A., Schilhabel,M., Schutte,B., Ganz,T., Linzmeier,R.,
            Ho,C.H., Hoang,B.V., McCray,P., Baumgart,C., Menzel,U.,
            Schattevoy,R. and Rosenthal,A.
TITLE       Direct Submission
JOURNAL     Submitted (22-FEB-2000) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 125544)
AUTHORS     Genome Sequencing Center Jena.
TITLE       Direct Submission
JOURNAL     Submitted (10-NOV-2000) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 125544)
AUTHORS     Genome Sequencing Center Jena.
TITLE       Direct Submission
JOURNAL     Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 125544)
AUTHORS     Lagemann,D. and Platzer,M.
TITLE       Direct Submission
JOURNAL     Submitted (02-APR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   6  (bases 1 to 125544)
AUTHORS     Lagemann,D. and Platzer,M.
TITLE       Direct Submission
JOURNAL     Submitted (06-JUL-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Apr 2, 2003 this sequence version replaced gi:21700555.
            -------------------- Genome Center
            Center: Institute of Molecular Biotechnology
            Center code: IMB
            Web site: http://genome.imb-jena.de/
            Contact: gscj-submit@genome.imb-jena.de
            -------------------- Project Information
            Center project name: H370
            Center clone name: Scb-561b17
            -------------------- Summary Statistics
            Sequencing vector: pUC18; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 127999 bases at least Q40
            Consensus quality: 128330 bases at least Q30
            Consensus quality: 128544 bases at least Q20
            Quality coverage: 11.05x
            --------------------
            This sequence was finished as follows unless otherwise noted; all
            regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.
            -------------------- Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            --------------------
FEATURES             Location/Qualifiers
     source          1..128544
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"

LOCUS       AF238378               128544 bp    DNA     linear   PRI 02-APR-2003
DEFINITION  Homo sapiens chromosome 8 clone SCb-561b17 map p22-p21, complete
            sequence.
ACCESSION   AF238378
VERSION     AF238378.5  GI:29469504
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128544)
AUTHORS     Polley,A., Baumgart,C., Blechschmidt,K., Dette,M.D., Jahn,N.,
            Menzel,U., Reichwald,K., Schilhabel,M.B., Schudy,A., Taudien,S.,
```

```
misc_feature    /map="p22-p21"
misc_feature    44529..44595
                /clone="SCb-561b17"
misc_feature    45703
                /note="single stranded/single chemistry region"
misc_feature    45815..45872
                /note="low quality region"
misc_feature    45824..45872
                /note="low quality region"
unsure          45843
                /note="single stranded/single chemistry region"
unsure          45847
unsure          45849
unsure          45858
unsure          45868
misc_feature    46235
misc_feature    47899..47900
                /note="low quality region"
unsure          47899..47900
                /note="low quality region"
misc_feature    49172
                /note="low quality region"
misc_feature    49220
unsure          49172
                /note="low quality region"
unsure          49339
                /note="low quality region"
misc_feature    49756..49807
                /note="low quality region"
misc_feature    49961
                /note="single stranded/single chemistry region"
misc_feature    52933..52948
                /note="low quality region"
misc_feature    52943
                /note="low quality region"
misc_feature    53544
unsure          53612..53926
misc_feature    53612..53926
                /note="low quality region"
misc_feature    53674..53677
                /note="single stranded/single chemistry region"
misc_feature    53779..53799
                /note="low quality region"
misc_feature    53787
                /note="low quality region"
misc_feature    54734..55129
misc_feature    55028..56165
                /note="single stranded/single chemistry region"
misc_feature    57952..58060
                /note="single stranded/single chemistry region"
misc_feature    57968..57974
                /note="single stranded/single chemistry region"
unsure          57968
unsure          57970
unsure          57974
misc_feature    58652
misc_feature    59561..59638
                /note="low quality region"
misc_feature    59925
                /note="single stranded/single chemistry region"
misc_feature    59929
unsure          61695..62103
                /note="low quality region"
misc_feature    61858..61859
                /note="low quality region"
misc_feature    62263..62485
                /note="single stranded/single chemistry region"
misc_feature    62655..63118
                /note="single stranded/single chemistry region"
misc_feature    62750..62785
                /note="low quality region"
```

```
unsure          62774..62777
unsure          62783
misc_feature    62801..62825
                /note="low quality region"
unsure          62819
unsure          62822
misc_feature    62872..62874
                /note="low quality region"
unsure          62908
misc_feature    62917..62987
                /note="low quality region"
misc_feature    62933..62935
                /note="low quality region"
unsure          62955
unsure          62957
misc_feature    63082
                /note="single clone coverage"
misc_feature    63110
                /note="low quality region"
misc_feature    63119..63281
                /note="low quality region"
unsure          63129..63131
misc_feature    63142..63258
                /note="single stranded/single chemistry region"
unsure          63159..63258
unsure          63197..63258
misc_feature    63281
misc_feature    63876..63878
                /note="low quality region"
misc_feature    63933..64021
                /note="low quality region"
misc_feature    63943
                /note="single stranded/single chemistry region"
misc_feature    63984..63985
                /note="low quality region"
misc_feature    64004..64007
                /note="low quality region"
misc_feature    70360
                /note="low quality region"
```

RESULT 6
AC021883/c

```
LOCUS       AC021883    232180 bp    DNA    linear    HTG    25-AUG-2000
DEFINITION  Homo sapiens chromosome 1 clone RP11-588H15, *** SEQUENCING IN
            PROGRESS ***, 40 unordered pieces.
ACCESSION   AC021883
VERSION     AC021883.4    GI:9910096
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 232180)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 232180)
```

US-10-045-180A-4 (1-19) x AF238378 (1-128544)

```
Qy         2 ArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63320 AGGACCCTCACCCTCCTCCTCTGCCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCA 63267
```

```
Alignment Scores:
Pred. No.:              0.00112      Length:        128544
Score:                  85.00        Matches:       18
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            94.44%       Indels:        0
DB:                                  Gaps:          0
```

```
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (21-JAN-2000) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
COMMENT   On Aug 25, 2000 this sequence version replaced gi:9838290.
          * ------- Genome Center -------
          * Center: Washington University Genome Sequencing Center
          * Center code: WUGSC
          * Web site: http://genome.wustl.edu/gsc/index.shtml
          * ------- Project Information -------
          * Center project name: H_NH0588H15
          * ------- Summary Statistics -------
          * Sequencing vector: M13; 86%
          * Sequencing vector: plasmid; 14%
          * Chemistry: Dye-primer ET; 86% of reads
          * Chemistry: Dye-terminator Big Dye; 14% of reads
          * Assembly program: Phrap; version 0.990319
          * Consensus quality: 211766 bases at least Q40
          * Consensus quality: 217990 bases at least Q30
          * Consensus quality: 221643 bases at least Q20
          * Insert size: 181000; agarose-fp
          * Insert size: 228280; sum-of-contigs
          * Quality coverage: 6.35 in Q20 bases; agarose-fp
          * Quality coverage: 4.69 in Q20 bases; sum-of-contigs
          *
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 40 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *        1      1312: contig of 1312 bp in length
          *     1313      1412: gap of unknown length
          *     1413      2446: contig of 1034 bp in length
          *     2447      2546: gap of unknown length
          *     2547      3878: contig of 1332 bp in length
          *     3879      3978: gap of unknown length
          *     3979      5215: contig of 1237 bp in length
          *     5216      5315: gap of unknown length
          *     5316      6544: contig of 1229 bp in length
          *     6545      6644: gap of unknown length
          *     6645      7940: contig of 1296 bp in length
          *     7941      8040: gap of unknown length
          *     8041      9190: contig of 1150 bp in length
          *     9191      9290: gap of unknown length
          *     9291     10909: contig of 1619 bp in length
          *    10910     11009: gap of unknown length
          *    11010     12203: contig of 1194 bp in length
          *    12204     12303: gap of unknown length
          *    12304     13542: contig of 1239 bp in length
          *    13543     13642: gap of unknown length
          *    13643     14982: contig of 1340 bp in length
          *    14983     15082: gap of unknown length
          *    15083     16905: contig of 1823 bp in length
          *    16906     17005: gap of unknown length
          *    17006     18051: contig of 1046 bp in length
          *    18052     18151: gap of unknown length
          *    18152     19373: contig of 1222 bp in length
          *    19374     19473: gap of unknown length
          *    19474     20668: contig of 1195 bp in length
          *    20669     20768: gap of unknown length
          *    20769     22399: contig of 1631 bp in length
          *    22400     22499: gap of unknown length
          *    22500     24200: contig of 1701 bp in length
          *    24201     24300: gap of unknown length
          *    24301     25667: contig of 1367 bp in length
          *    25668     25767: gap of unknown length
          *    25768     27705: contig of 1938 bp in length
          *    27706     27805: gap of unknown length
          *    27806     29643: contig of 1838 bp in length
          *    29644     29743: gap of unknown length
          *    29744     31718: contig of 1975 bp in length
          *    31719     31818: gap of unknown length
          *    31819     33299: contig of 1481 bp in length
          *    33300     33399: gap of unknown length
          *    33400     35588: contig of 2189 bp in length
          *    35589     35688: gap of unknown length
          *    35689     37968: contig of 2280 bp in length
          *    37969     38068: gap of unknown length
          *    38069     40341: contig of 2273 bp in length
          *    40342     40441: gap of unknown length
          *    40442     42574: contig of 2133 bp in length
          *    42575     42674: gap of unknown length
          *    42675     44938: contig of 2264 bp in length
          *    44939     45038: gap of unknown length
          *    45039     50126: contig of 5088 bp in length
          *    50127     50226: gap of unknown length
          *    50227     52594: contig of 2368 bp in length
          *    52595     52694: gap of unknown length
          *    52695     57156: contig of 4462 bp in length
          *    57157     57256: gap of unknown length
          *    57257     61310: contig of 4054 bp in length
          *    61311     61410: gap of unknown length
          *    61411     65807: contig of 4397 bp in length
          *    65808     65907: gap of unknown length
          *    65908     70160: contig of 4253 bp in length
          *    70161     70260: gap of unknown length
          *    70261     76919: contig of 6659 bp in length
          *    76920     77019: gap of unknown length
          *    77020     85764: contig of 8745 bp in length
          *    85765     85864: gap of unknown length
          *    85865    106630: contig of 20766 bp in length
          *   106631    106730: gap of unknown length
          *   106731    134020: contig of 27290 bp in length
          *   134021    134120: gap of unknown length
          *   134121    165082: contig of 30962 bp in length
          *   165083    165182: gap of unknown length
          *   165183    196240: contig of 31058 bp in length
          *   196241    196340: gap of unknown length
          *   196341    232180: contig of 35840 bp in length
FEATURES             Location/Qualifiers
     source          1..232180
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-588H15"
     misc_feature    1..1312
                     /note="assembly_name:Contig14"
     misc_feature    1413..2446
                     /note="assembly_name:Contig15"
     misc_feature    2547..3878
                     /note="assembly_name:Contig30"
     misc_feature    3979..5215
                     /note="assembly_name:Contig35"
     misc_feature    5316..6544
                     /note="assembly_name:Contig40"
     misc_feature    6645..7940
                     /note="assembly_name:Contig42"
     misc_feature    8041..9190
                     /note="assembly_name:Contig44"
     misc_feature    9291..10909
                     /note="assembly_name:Contig47"
     misc_feature    11010..12203
                     /note="assembly_name:Contig48"
     misc_feature    12304..13542
                     /note="assembly_name:Contig52"
     misc_feature    13643..14982
                     /note="assembly_name:Contig53"
     misc_feature    15083..16905
                     /note="assembly_name:Contig54"
     misc_feature    17006..18051
```

```
     misc_feature    18152..19373
                      /note="assembly_name:Contig55"
     misc_feature    19474..20668
                      /note="assembly_name:Contig56"
     misc_feature    20769..22399
                      /note="assembly_name:Contig57"
     misc_feature    22500..24200
                      /note="assembly_name:Contig58"
     misc_feature    24301..25667
                      /note="assembly_name:Contig59"
     misc_feature    25768..27705
                      /note="assembly_name:Contig60"
     misc_feature    27806..29643
                      /note="assembly_name:Contig61"
     misc_feature    29744..31718
                      /note="assembly_name:Contig62"
     misc_feature    31819..33299
                      /note="assembly_name:Contig64"
     misc_feature    33400..35588
                      /note="assembly_name:Contig65"
     misc_feature    35689..37968
                      /note="assembly_name:Contig66"
     misc_feature    38069..40341
                      /note="assembly_name:Contig67"
     misc_feature    40442..42574
                      /note="assembly_name:Contig68"
     misc_feature    42675..44938
                      /note="assembly_name:Contig69"
     misc_feature    45039..50126
                      /note="assembly_name:Contig70"
     misc_feature    50227..52594
                      /note="assembly_name:Contig71"
     misc_feature    52695..57156
                      /note="assembly_name:Contig72"
     misc_feature    57257..61310
                      /note="assembly_name:Contig73"
     misc_feature    /note="assembly_name:Contig74
                      clone_end:SP6
                      vector_side:right"

Alignment Scores:
Pred. No.:              9.75        Length:         232180
Score:                  65.00       Matches:        15
Percent Similarity:     89.47%      Conservative:   2
Best Local Similarity:  78.95%      Mismatches:     2
Query Match:            72.22%      Indels:         0
                                    Gaps:           0

Qy      1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
            |||||||||||||||||::|||::||||||||
Db  14901 ATGAGAACCCTCACCATCCTCACTGCGTGTTCTCCGTGGCCCTCCAGGCACAGGCT 14845

US-10-045-180A-4 (1-19) x AC021883 (1-232180)

RESULT 7
LOCUS       AX405993       398 bp    DNA    linear    PAT 14-JUN-2002
DEFINITION  Sequence 408 from Patent WO0222660.
ACCESSION   AX405993
VERSION     AX405993.1  GI:21439425
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tang,Y.T., Liu,C., Zhou,P., Asundi,V., Zhang,J., Zhao,Q.A., Ren,F.,
            Xue,A.J., Yang,Y., Wehrman,T. and Drmanac,R.T.
  TITLE     Novel nucleic acids and polypeptides
  JOURNAL   Patent: WO 0222660-A 408 21-MAR-2002;
            HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1..398
```

```
     CDS             /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     30..281
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD35100.1"
                     /db_xref="GI:21439426"
                     /translation="MRTLTLTFAVLLVALQAKAEPLQAEDDPLQAKAYEADAQEQRGA
                     NDQDFAVSFAEDASSLRALGKRHQHCRARSVERKTKHF"

BASE COUNT      102 a      108 c       93 g       95 t
ORIGIN

US-10-045-180A-4 (1-19) x AX405993 (1-398)

RESULT 8
LOCUS       AR270635       452 bp    DNA    linear    PAT 10-APR-2003
DEFINITION  Sequence 1198 from patent US 6500938.
ACCESSION   AR270635
VERSION     AR270635.1  GI:29701869
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Au-Young,J. and Seilhamer,J.J.
  TITLE     Composition for the detection of signaling pathway gene expression
  JOURNAL   Patent: US 6500938-A 1198 31-DEC-2002;
FEATURES             Location/Qualifiers
     source          1..452
                     /organism="unknown"

BASE COUNT      124 a      118 c       98 g      112 t
ORIGIN

Alignment Scores:
Pred. No.:              0.0187      Length:         398
Score:                  64.00       Matches:        15
Percent Similarity:     89.47%      Conservative:   2
Best Local Similarity:  78.95%      Mismatches:     2
Query Match:            71.11%      Indels:         0
                                    Gaps:           0

Qy      1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
            |||||||||||||||||::|||::||||||||
Db     30 ATGAGAACCCTCACCATCCTCACTGCGTGTTCTCCGTGGCCCTCCAGGCCAAGGCT 86

US-10-045-180a-4 (1-19) x AR270635 (1-452)

RESULT 9
LOCUS       I49588         452 bp    DNA    linear    PAT 07-OCT-1997
DEFINITION  Sequence 6 from patent US 5641497.
ACCESSION   I49588
VERSION     I49588.1  GI:2471808
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Bevins,C.L. and Jones,D.E.

Alignment Scores:
Pred. No.:              0.0213      Length:         452
Score:                  64.00       Matches:        15
Percent Similarity:     89.47%      Conservative:   2
Best Local Similarity:  78.95%      Mismatches:     2
Query Match:            71.11%      Indels:         0
                                    Gaps:           0

Qy      1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
            |||||||||||||||||::|||::||||||||
Db     19 ATGAGAACCCTCACCATCCTCACTGCGTGTTCTCCGTGGCCCTCCAGGCCAAGGCT 75
```

TITLE   Gastrointestinal defensins, cDNA sequences and method for the
            production and use thereof
JOURNAL   Patent: US 5641497-A 6 24-JUN-1997;
FEATURES   Location/Qualifiers
     source   1. .452
                     /organism="unknown"
BASE COUNT   124 a   118 c   98 g   112 t
ORIGIN

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaThrP2Ala 19
              |||||||||||||||||::||::||| ||||||||||||||||||||||||
Db   19 ATGAGAACCCTCACCATCCTCACTGCTGTTTCCTCCGTGGCCCTTCCAGGCCAAGGCT 75

US-10-045-180A-4 (1-19) x 149588 (1-452)

Alignment Scores:
Pred. No.:   0.0213   Length: 452
Score: 64.00   Matches: 15
Percent Similarity: 89.47%   Conservative: 2
Best Local Similarity: 78.95%   Mismatches: 2
Query Match: 71.11%   Indels: 0
DB: 6   Gaps: 0

RESULT 10
HUMDFSNSIX
LOCUS   HUMDFSNSIX   452 bp   mRNA   linear   PRI 27-APR-1993
DEFINITION Homo sapiens defensin 6 mRNA, complete cds.
ACCESSION   M98331
VERSION   M98331.1 GI:181546
KEYWORDS   defensin 6.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 452)
  AUTHORS   Jones,D.E. and Bevins,C.L.
  TITLE   Defensin-6 mRNA in human Paneth cells: implications for
            antimicrobial peptides in host defense of the human bowel
  JOURNAL   FEBS Lett. 315 (2), 187-192 (1993)
  MEDLINE   93114459
  PUBMED   8417977
COMMENT   Original source text: Homo sapiens small intestine cDNA to mRNA.
FEATURES   Location/Qualifiers
     source   1. .452
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="small intestine"
     5'UTR   1. .18
     CDS   19. .321
                     /codon_start=1
                     /note="preprepeptide"
                     /product="defensin 6"
                     /protein_id="AAB59357.1"
                     /db_xref="GI:181547"
                     /translation="MRTLTLLTAVLLVALVALQAKAEPLQAEDDPLQAKAYEADAQEQRGA
                     NDQDFAVSFAEDASSSLRALGSSTRAFTCHCRRSCYSTEYSYGTCTVMGINHRFCCL"
     3'UTR   319. .440
     variation   390
                     /replace="t"
     polyA_signal   424. .429

ORIGIN
BASE COUNT   124 a   118 c   98 g   112 t

US-10-045-180A-4 (1-19) x HUMDFSNSIX (1-452)

Alignment Scores:
Pred. No.:   0.0213   Length: 452
Score: 64.00   Matches: 15
Percent Similarity: 89.47%   Conservative: 2
Best Local Similarity: 78.95%   Mismatches: 2
Query Match: 71.11%   Indels: 0
DB: 6   Gaps: 0

RESULT 11
HSU33317
LOCUS   HSU33317   3060 bp   DNA   linear   PRI 23-FEB-1996
DEFINITION Human defensin 6 (HD-6) gene, complete cds.
ACCESSION   U33317
VERSION   U33317.1 GI:1200181
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3060)
  AUTHORS   Mallow,E.B., Harris,A., Salzman,N., Russell,J.P.,
            DeBerardinis,R.J., Ruchelli,E. and Bevins,C.L.
  TITLE   Human enteric defensins. Gene structure and developmental
            expression
  JOURNAL   J. Biol. Chem. 271 (8), 4038-4045 (1996)
  MEDLINE   96223969
  PUBMED   8626737
REFERENCE   2 (bases 1 to 3060)
  AUTHORS   Charles L. Bevins.
  TITLE   Direct Submission
  JOURNAL   Submitted (07-AUG-1995) Charles L. Bevins, Pediatrics, The
            Children's Hospital of Philadelphia, 34th and Civic Center Blvd.,
            Philadelphia, PA 19104, USA
FEATURES   Location/Qualifiers
     source   1. .3060
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CAAT_signal   1277
     TATA_signal   1341
     gene   1370. .2730
                     /gene="HD-6"
     mRNA   join(1370. .1605,2517. .2730)
                     /gene="HD-6"
     5'UTR   1370. .1410
                     /gene="HD-6"
     CDS   join(1411. .1605,2517. .2627)
                     /gene="HD-6"
                     /codon_start=1
                     /product="defensin 6"
                     /protein_id="AAC50382.1"
                     /db_xref="GI:1200182"
                     /translation="MRTLTLLTAVLLVALVALQAKAEPLQAEDDPLQAKAYEADAQEQRGA
                     NDQDFAVSFAEDASSSLRALGSSTRAFTCHCRRSCYSTEYSYGTCTVMGINHRFCCL"
     3'UTR   2628. .2730
                     /gene="HD-6"
     polyA_signal   2730
                     /gene="HD-6"
ORIGIN
BASE COUNT   892 a   678 c   612 g   878 t

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaThrPAla 19
              |||||||||||||||||::||::||| ||||||||||||||||||||||||
Db   19 ATGAGAACCCTCACCATCCTCACTGCTGTTTCCTCCGTGGCCCTTCCAGGCCAAGGCT 75

US-10-045-180A-4 (1-19) x HSU33317 (1-3060)

Alignment Scores:
Pred. No.:   0.158   Length: 3060
Score: 64.00   Matches: 15
Percent Similarity: 89.47%   Conservative: 2
Best Local Similarity: 78.95%   Mismatches: 2
Query Match: 71.11%   Indels: 0
DB: 9   Gaps: 0

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaThrPAla 19
              |||||||||||||||||::||::||| ||||||||||||||||||||||||
Db   1411 ATGAGAACCCTCACCATCCTCACTGCTGTTTCCTCCGTGGCCCTTCCAGGCCAAGGCT 1467

RESULT 12

```
LOCUS       AF314060               79730 bp    DNA     linear   PRI 09-JUL-2002
DEFINITION  Homo sapiens chromosome 8 clone Qu-L-i2046 map 8p22-p21, complete
            sequence.
ACCESSION   AF314060
VERSION     AF314060.2  GI:21717333
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 79730)
  AUTHORS   Polley,A., Baumgart,C., Blechschmidt,K., Dette,M.D., Jahn,N.,
            Menzel,U., Reichwald,K., Schilhabel,M.B., Schudy,A., Taudien,S.,
            Wen,G., Schutte,B., Malik,M., Peng,J.Hong., McCray,P. and
            Rosenthal,A.
  TITLE     Chromosome 8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 79730)
  AUTHORS   Genome Sequencing Center Jena.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2000) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 79730)
  AUTHORS   Lagemann,D. and Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
  COMMENT   On Jul 9, 2002 this sequence version replaced gi:10998844.
            --------------------- Genome Center
            Center: Institute of Molecular Biotechnology
            Center code: IMB
            Web site: http://genome.imb-jena.de/
            Contact: gscj-submit@genome.imb-jena.de
            --------------------- Project Information
            Center project name: H376
            Center clone name: Qu-L-i2046
            --------------------- Summary Statistics
            Sequencing vector: pUC18; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 79710 bases at least Q40
            Consensus quality: 79727 bases at least Q30
            Consensus quality: 79729 bases at least Q20
            Quality coverage: 12.19x
            ---------------------
            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one pUC18 subclone; and the
            assembly was confirmed by restriction digest.
            ---------------------
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
FEATURES             Location/Qualifiers
     source          1..79730
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8p22-p21, complete sequence."
                     /clone="Qu-L-i2046"
     variation       172
     variation       500
                     /note="C substituted in clone: GS1-P1-861"
                     /replace="A"
     variation       679
                     /note="T substituted in clone: SCb-633e22, Qu-L-i2046"
                     /replace="C"
     variation       729
                     /note="G substituted in clone: SCb-633e22"
                     /replace="A"
     variation       791
                     /note="T substituted in clone: Qu-L-i2046"
                     /replace="G"
     variation       1245
                     /note="A substituted in clone: Qu-L-i2046, GS1-P1-861,
                     SCb-633e22"
                     /replace="G"
     variation       1668
                     /note="C substituted in clone: SCb-633e22"
                     /note="Qu-L-i2046"
                     /replace="T"
     variation       1777
                     /note="A substituted in clone: Qu-L-i2046, SCb-633e22,
                     GS1-P1-861"
                     /replace="G"
     variation       1866
                     /note="G substituted in clone: GS1-P1-861"
                     /replace="T"
     variation       2621
                     /note="A substituted in clone: Qu-L-i2046, SCb-633e22"
                     /replace="G"
     variation       3021..3022
                     /note="C substituted in clone: GS1-P1-861"
                     /replace="T"
     variation       3284
                     /note="deleted in clone: SCb-633e22, GS1-P1-861"
                     /replace=""
     variation       3894
                     /note="G substituted in clone: SCb-633e22"
                     /replace="A"
     variation       4041
                     /note="A substituted in clone: SCb-633e22"
                     /replace="T"
     variation       4204
                     /note="A substituted in clone: GS1-P1-861"
                     /replace="C"
     variation       4510
                     /note="C substituted in clone: GS1-P1-861, SCb-633e22,
                     CTD-2629116"
                     /note="deleted in clone: GS1-P1-861, SCb-633e22,
                     /replace="T"
     variation       4808
                     /note="T substituted in clone: Qu-L-i2046, SCb-633e22"
                     /replace="C"
     variation       4949
                     /note="A substituted in clone: Qu-L-i2046, SCb-633e22"
                     /replace="T"
     variation       5121
                     /note="A substituted in clone: Qu-L-i2046, SCb-633e22"
                     /replace="G"
     variation       5142
                     /note="T substituted in clone: Qu-L-i2046, SCb-633e22"
                     /replace="C"
     variation       6261
                     /note="A substituted in clone: GS1-P1-861"
                     /replace="G"
     variation       6473
                     /note="T substituted in clone: Qu-L-i2046, SCb-633e22"
                     /replace="C"
     variation       6870
                     /note="T substituted in clone: SCb-633e22"
                     /replace="G"
     variation       7648
                     /note="C substituted in clone: GS1-P1-861"
                     /replace="T"
```

```
LOCUS       AF233439              150940 bp    DNA     linear   PRI 01-APR-2003
DEFINITION  Homo sapiens chromosome 8 clone CTD-2629116 map p22-p21, complete
            sequence.
ACCESSION   AF233439
VERSION     AF233439.6  GI:29423859
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150940)
  AUTHORS   Reichwald,K., Menzel,U., Dette,M., Baumgart,C., Jahn,N., Wen,G.,
            Schudy,A., Schilhabel,M. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2000) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   2  (bases 1 to 150940)
  AUTHORS   Schattevoy,R. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 150940)
  AUTHORS   Genome Sequencing Center Jena.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 150940)
  AUTHORS   Reichwald,K. and Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2001) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena, Thuringia 07745, Germany
REFERENCE   5  (bases 1 to 150940)
  AUTHORS   Lagemann,D. and Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   6  (bases 1 to 150940)
  AUTHORS   Lagemann,D. and Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   7  (bases 1 to 150940)
  AUTHORS   Lagemann,D. and Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     On Apr 1, 2003 this sequence version replaced gi:29336172.
            ----------------------- Genome Center
               Center: Institute of Molecular Biotechnology
               Center code: IMB
               Web site: http://genome.imb-jena.de/
               Contact: gscj-submit@genome.imb-jena.de
            ----------------------- Project Information
               Center project name: H350
               Center clone name: CTD-2629116
            ----------------------- Summary Statistics
               Sequencing vector: pUC18; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 150881 bases at least Q40
               Consensus quality: 150926 bases at least Q30
               Consensus quality: 150940 bases at least Q20
               Quality coverage: 22.17x
            -----------------------
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            -----------------------
            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.
```

variation 7706
           /note="T substituted in clone: CTD-2629116"
           /replace="A"
variation 7706
           /note="G substituted in clone: GS1-P1-861"
           /replace="A"
variation 7767
           /note="C substituted in clone: GS1-P1-861"
           /replace="T"
variation 7828..7829
           /note="C substituted in clone: Qu-L-i2046, SCb-633e22"
           /replace="A"
variation 7841
           /note="AA substituted in clone: SCb-633e22, Qu-L-i2046"
           /replace="GG"
variation 7957
           /note="deleted in clone: GS1-P1-861"
           /replace=""
variation 8629
           /note="T substituted in clone: Qu-L-i2046, SCb-633e22"
           /replace="G"
variation 8676
           /note="A substituted in clone: SCb-633e22, Qu-L-i2046"
           /replace="G"
variation 9188
           /note="C substituted in clone: GS1-P1-861"
           /replace="T"
variation 9784
           /note="deleted in clone: SCb-633e22"
           /replace=""
variation Qu-L-i2046"
           /note="A substituted in clone: GS1-P1-861, SCb-633e22,
           /replace="C"
variation 1526
           /note="T substituted in clone: SCb-633e22"
           /replace="C"
variation 10659
           /note="A substituted in clone: Qu-L-i2046, SCb-633e22"
           /replace="C"
variation 10864
           /note="A substituted in clone: SCb-633e22"
           /replace="G"
variation 10903..10905
           /note="deleted in clone: GS1-P1-861, CTD-2629116,
           /replace=""
variation 11036
           /note="C substituted in clone: SCb-633e22, GS1-P1-861,
           Qu-L-i2046"
           /replace="A"
variation 11087
           /note="T substituted in clone: SCb-633e22, GS1-P1-861,
           /replace="C"
variation 11163
           /note="C substituted in clone: SCb-633e22"
           /replace="A"
variation 11726
           /note="G substituted in clone: SCb-633e22, GS1-P1-861"

```
US-10-045-180A-4 (1-19) x AF314060 (1-79730)

Alignment Scores:
Pred. No.:                                                      4.85
Score:                                                         64.00
Percent Similarity:                89.47%     Length:           79730
Best Local Similarity:             78.95%     Matches:             15
Query Match:                       71.11%     Mismatches:           2
                                              Conservative:         2
                                              Indels:               0
                                              Gaps:                 0

DB:      AF233439/c
Qy       1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
           |||||||||||||||||||:::||||||||::|:||||
Db   39728 ATGAGAACCCTCACCATCCTCACTGCGTTCTTCCTCGTGGCCCTCCAGGCCAAGGCT 39784
```

All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

```
FEATURES             Location/Qualifiers
     source          1..150940
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="p22-p21"
                     /clone="CTD-2629116"
     misc_feature    31743..32232
                     /note="low quality region"
     misc_feature    38626..38692
                     /note="single stranded/single chemistry region"
     misc_feature    93307..93310
                     /note="single stranded/single chemistry region
                     low quality region"
     misc_feature    93393..93395
                     /note="low quality region"
     misc_feature    93427..93428
                     /note="low quality region"
     misc_feature    93434..93436
                     /note="low quality region"
     unsure          93435
                     /note="low quality region"
     misc_feature    93481
                     /note="low quality region"
     unsure          93481
                     /note="low quality region"
     misc_feature    93534
                     /note="low quality region"
BASE COUNT     39368 a   33223 c   36391 g   41958 t
ORIGIN
```

US-10-045-180A-4 (1-19) x AF233439 (1-150940)

```
Qy           1 MetArgThrIleuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
               |||||||||||||||:::|||||:||:|||
Db       85970 ATGAGAACCCTCACCATCCTGTCTGCTTTCCTGCTGGTGGCCCTCCAGGCCAAGGCT 85914
```

Alignment Scores:

Pred. No.:	9.47	Length:	150940
Score:	64.00	Matches:	15
Percent Similarity:	89.47%	Conservative:	2
Best Local Similarity:	78.95%	Mismatches:	2
Query Match:	71.11%	Indel:	0
DB:	9	Gaps:	0

RESULT 14

```
LOCUS       AF200455/c          158653 bp    DNA     linear   PRI 09-JAN-2003
            AF200455/c
DEFINITION  Homo sapiens chromosome 8 clone SCb-540n10 map p22-p21, complete
            sequence.
ACCESSION   AF200455
VERSION     AF200455.3  GI:27552755
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158653)
  AUTHORS   Schudy,A., Blechschmidt,K., Menzel,U., Polley,A., Reichwald,K.,
            Rump,A., Schilhabel,M.B., Taudien,S., Wen,G., Schutte,B., Malik,M.,
            Peng,J.Hong., McCray,P. and Rosenthal,A.
  TITLE     Chromosome 8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158653)
  AUTHORS   Schudy,A., Schilhabel,M., Schutte,B., Malik,M., Hong Peng,J.,
            McCray,P., Baumgart,C., Menzel,U., Weber,J., Schattevoy,R. and
            Rosenthal,A.
```

TITLE   Direct Submission
JOURNAL   Submitted (28-OCT-1999) Genome Analysis, Institute of Molecular
          Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   3 (bases 1 to 158653)
  AUTHORS   Genome Sequencing Center Jena.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE   4 (bases 1 to 158653)
  AUTHORS   Lagemann,D. and Platzer,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-JAN-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
          On Jan 9, 2003 this sequence version replaced gi:8151988.
COMMENT   Genome Center
            Center: Insitute of Molecular Biotechnology
            Center code: IMB
            Web site: http://genome.imb-jena.de/
            Contact: gscj-submit@genome.imb-jena.de
            -------------- Project Information
            Center project name: Scb-540n10
            Center clone name: H293
            -------------- Summary Statistics
            Sequencing vector: pUC18; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; Version 0.990329
            Consensus quality: 157265 bases at least Q40
            Consensus quality: 158014 bases at least Q30
            Consensus quality: 158652 bases at least Q20
            Quality coverage: 8.15x

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.

            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.

```
FEATURES             Location/Qualifiers
     source          1..158653
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="p22-p21"
                     /clone="SCb-540n10"
     misc_feature    773..805
                     /note="single stranded/single chemistry region"
     misc_feature    1550..1558
                     /note="low quality region"
     unsure          1554
                     /note="low quality region"
     misc_feature    1557..1596
                     /note="single stranded/single chemistry region"
     misc_feature    1571
                     /note="low quality region"
     misc_feature    2687..3061
                     /note="single stranded/single chemistry region"
     misc_feature    14120..14203
                     /note="single stranded/single chemistry region"
     unsure          25023
     unsure          25552
     misc_feature    25774
```

misc_feature   26465..26484   /note="low quality region"
misc_feature   26485..26616   /note="single stranded/single chemistry region"
misc_feature   26592   /note="single clone coverage"
unsure   32723
unsure   38261
unsure   42450..42453
misc_feature   43111..43237   /note="single stranded/single chemistry region"
unsure   43134
unsure   43160
unsure   43183
misc_feature   43191   /note="low quality region"
unsure   46094..46149
misc_feature   46189   /note="single stranded/single chemistry region"
unsure   46238..46240
unsure   48224..48285
misc_feature   53580..53612   /note="single stranded/single chemistry region"
misc_feature   53587   /note="single stranded/single chemistry region"
misc_feature   53589   /note="low quality region"
misc_feature   53595..53597   /note="low quality region"
misc_feature   53595..53597   /note="low quality region"
misc_feature   53613..53709   /note="single chemistry region"
misc_feature   53641..53645   /note="single clone coverage"
misc_feature   53649..53655   /note="low quality region"
unsure   53649..53655
unsure   53696..53709
unsure   53699
misc_feature   53788   /note="low quality region"
misc_feature   53791   /note="low quality region"
unsure   53816..53822   /note="low quality region"
misc_feature   53660..53661   /note="low quality region"
misc_feature   53666..53675   /note="low quality region"
misc_feature   53672   /note="low quality region"
unsure   53688
misc_feature   53696..53709
unsure   53649..53655
misc_feature   53641..53645   /note="single clone coverage"
misc_feature   53613..53709
misc_feature   53595..53597   /note="low quality region"
misc_feature   53595..53597   /note="low quality region"
misc_feature   53589   /note="low quality region"
misc_feature   53587   /note="single stranded/single chemistry region"
misc_feature   53580..53612   /note="single stranded/single chemistry region"
unsure   48224..48285
unsure   46238..46240
misc_feature   46189   /note="single stranded/single chemistry region"
unsure   46094..46149
misc_feature   43191   /note="low quality region"
unsure   43183
unsure   43160
unsure   43134
misc_feature   43111..43237   /note="single stranded/single chemistry region"
unsure   42450..42453
unsure   38261
unsure   32723
misc_feature   26592   /note="single clone coverage"
misc_feature   26485..26616   /note="single stranded/single chemistry region"
misc_feature   26465..26484   /note="low quality region"

misc_feature   63929..63985   /note="single stranded/single chemistry region"
misc_feature   63927..64074   /note="single stranded/single chemistry region"
misc_feature   63916..63917   /note="low quality region"
misc_feature   63898..63903   /note="low quality region"
misc_feature   63882..63889   /note="low quality region"
misc_feature   63873..63877   /note="low quality region"
unsure   63860
misc_feature   63860   /note="low quality region"
misc_feature   63854   /note="low quality region"
misc_feature   63842   /note="low quality region"
misc_feature   63836   /note="low quality region"
misc_feature   63816..63822   /note="low quality region"
misc_feature   63791   /note="low quality region"
misc_feature   63788   /note="low quality region"
unsure   63699
unsure   63649..53709
misc_feature   63660..63661   /note="low quality region"
misc_feature   63651
misc_feature   63672
unsure   63688

unsure   63958   /note="low quality region"
unsure   63961
unsure   63982
misc_feature   64015..64017   /note="low quality region"
misc_feature   64030..64036   /note="low quality region"
misc_feature   64045..64048   /note="low quality region"
misc_feature   73345..73421   /note="low quality region"
misc_feature   73121..83206   /note="single clone coverage"
misc_feature   75927..76057   /note="single stranded/single chemistry region"
misc_feature   83076..83120   /note="single stranded/single chemistry region"
misc_feature   83099..83182   /note="single stranded/single chemistry region"
unsure   83123
unsure   83143
misc_feature   83147
misc_feature   83207..83309   /note="single stranded/single chemistry region"
misc_feature   83243..83245   /note="low quality region"
misc_feature   84552   /note="low quality region"
misc_feature   84556..84700   /note="single stranded/single chemistry region"
misc_feature   84589..84590   /note="single chemistry region"
misc_feature   84607..84610   /note="low quality region"

Alignment Scores:

Pred. No.:	9.97
Score:	64.00
Percent Similarity:	89.47%
Best Local Similarity:	78.95%
Query Match:	71.11%
DB:	9

Length:	158653
Matches:	15
Conservative:	2
Mismatches:	2
Indels:	0
Gaps:	0

US-10-045-180A-4 (1-19) x AF200455 (1-158653)

```
Qy      1 MetArgThrLeuThrLeuLeuSerAlaLaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          ||||||||||||||||||||||:::||||||::|||||||||||||||||||||||||| |||
Db  29175 ATGAGAGRACCCTCACCTCACTCGTGTTCTCCTCGTGGCCCTCCPAGGCCAAGGCT 29119
```

RESULT 15
AC079018/c
LOCUS   AC079018   168702 bp   DNA   linear   PRI 15-AUG-2002
DEFINITION   Homo sapiens chromosome 8, clone RP11-161B1, complete sequence.
ACCESSION   AC079018
VERSION   AC079018.10   GI:22263526
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 168702)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,B.
  TITLE   Homo sapiens chromosome 8, clone RP11-161B1
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 168702)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
             Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
             Chepel,Y., Colangelo,M., Collins,F., Collymore,A., Cooke,P.,
             DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
             FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
             Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
             Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,

REFERENCE 3 (bases 1 to 168702)
AUTHORS   Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
           Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
           McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
           Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
           O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K.,
           Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
           Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
           Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
           Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
           Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
           Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
           Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (15-AUG-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 168702)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
           Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
           Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
           Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
           Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
           Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
           Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
           Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
           Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
           McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
           Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
           O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
           Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
           Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
           Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
           Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
           Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
           Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Aug 15, 2002 this sequence version replaced gi:22123231.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ------- Genome Center
           Center: Whitehead Institute/MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu
           ------- Project Information
           Center project name: L10753
           Center clone name: 161_B_1

FEATURES   Location/Qualifiers
     source   1..168702
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone=RP11-161B1"
                     /clone_lib="RPCI-11 Human Male BAC"
                     /note="PCR product sequence only"
     repeat_region   1..515
     unsure   355..471
     repeat_region   528..631
                     /rpt_family="(TG)n"
     unsure   832..836
                     /note="30 qual SNGL region"
     repeat_region   2098..2371
                     /rpt_family="AluJo"
     repeat_region   2909..3311
                     /rpt_family="L2"
     repeat_region   3394..3422
                     /rpt_family="Arr_rich"
                     complement(3799..4104)
                     /rpt_family="AluSx"
                     complement(4115..4425)
                     /rpt_family="AluY"
     repeat_region   4907..4934
                     /rpt_family="AluY"
     repeat_region   6023..6326
                     /rpt_family="(TTTTC)n"
     repeat_region   7266..7764
                     /rpt_family="AluY"
     repeat_region   7767..8016
                     /rpt_family="MER74B"
     repeat_region   10411..10478
                     /rpt_family="AluSx"
     repeat_region   10482..10962
                     complement(10482..10962)
                     /rpt_family="(TA)n"
     repeat_region   11185
                     /rpt_family="MER31B"
     complement(10984..11185)
                     /rpt_family="MSTA"
     repeat_region   11218..11522
                     /rpt_family="AluSx"
     repeat_region   11733..11851
                     /rpt_family="L1M4"
     complement(11919..12193)
                     /rpt_family="AluSx"
     repeat_region   12434..12735
                     /rpt_family="AluSx"
     repeat_region   12761..13748
                     /rpt_family="AluY"
     complement(13875..14169)
                     /rpt_family="L1M4"
     repeat_region   14995..15035
                     /rpt_family="AluSx"
     repeat_region   15872..16053
                     /rpt_family="(TGAA)n"
     repeat_region   17044..17211
                     /rpt_family="L2"
     repeat_region   17301..17598
                     /rpt_family="MLT1H"
     complement(18497..18646)
                     /rpt_family="MIR"
     complement(19356..24412)
                     /rpt_family="L1PA5"
     complement(24410..24506)
                     /rpt_family="L1PA5"
     repeat_region   24985..25170
                     /rpt_family="(CA)n"
     repeat_region   25202..25467
                     /rpt_family="(CA)n"
     repeat_region   25486..25588
                     /rpt_family="(CA)n"
     repeat_region   25613..25784
                     /rpt_family="(CA)n"

repeat_region 25831. .25969
               /rpt_family="(CA)n"
repeat_region 25949. .26079
               /rpt_family="(CATA)n"
repeat_region 26481. .26856
               /rpt_family="(CA)n"
repeat_region complement(27532. .27905)
               /rpt_family="MER115"
repeat_region complement(27906. .28004)
               /rpt_family="MER4A"
repeat_region complement(28176. .28667)
               /rpt_family="MER4A"
repeat_region 29190. .29439
               /rpt_family="MER74A"
repeat_region complement(29440. .29898)
               /rpt_family="L2"
repeat_region 29899. .29971
               /rpt_family="MLT1C"
repeat_region 30095. .30125
               /rpt_family="L2"
repeat_region 30647. .30668
               /rpt_family="(GAATG)n"
repeat_region complement(31093. .31397)
               /rpt_family="AT_rich"
repeat_region 31444. .31570
               /rpt_family="AluSp"
repeat_region 31821. .32123
               /rpt_family="LIM4"
repeat_region 32692. .32984
               /rpt_family="AluSx"
repeat_region complement(33681. .33913)
               /rpt_family="AluSc"

Alignment Scores:
Pred. No.:   Length: 168702
Score: 10.6   Matches: 15
Percent Similarity: 64.00   Conservative: 2
Best Local Similarity: 89.47%   Mismatches: 2
Query Match: 78.95%   Indels: 0
DB: 71.11%   Gaps: 0
                        9

US-10-045-180A-4 (1-19) x AC079018 (1-168702)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
             |||||||||||||::|||:::|||||||||::||||||||||||||||||| |||
Db   100083 ATGAGAACCCTCACCATCCTCACTGCTGTTCTCCTCGTTGGCCCCTTCCAGGCCAAGGCT 100027

Search completed: December 17, 2003, 16:37:39
Job time : 603.622 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 14:50:33

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Title: US-10-045-180A-4
Perfect Score: 90
Sequence: 1 MRTLTLLSAFLIVALQAWA 19

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

(without alignments)
Search time 664.293 Seconds
695.153 Million cell updates/sec

Command line parameters:
-MODEL=frame+_p2n.model
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145024_12540/app_query.fasta_1.860
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10045180 @CGN_1_6100 @runat_17122003_145024_12540 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: em_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estom:*
16: em_estfun:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	63	70.0	221	10	BF895944	BF895944 CM2-MT015
c	2	61	67.8	168	12	BI023130	BI023130 CM0-MT039
c	3	61	67.8	194	10	BG193795	BG193795 RST12931
c	4	61	67.8	243	13	BQ334320	BQ334320 QV3-MT012
c	5	61	67.8	251	9	AA321190	AA321190 EST23992
c	6	61	67.8	281	9	AA321281	AA321281 EST23811
c	7	61	67.8	289	9	AA321317	AA321317 EST23784
c	8	61	67.8	290	10	BF999113	BF999113 QV3-GN010
c	9	61	67.8	294	9	AA321316	AA321316 EST23783
c	10	61	67.8	298	9	AA321377	AA321377 EST23847
	11	61	67.8	312	9	AA321102	AA321102 EST23896
	12	61	67.8	320	9	AA321426	AA321426 EST23923
	13	61	67.8	327	9	AA321192	AA321192 EST23994
	14	61	67.8	348	10	BF893385	BF893385 QV3-MT012
c	15	61	67.8	349	12	BI021617	BI021617 CM3-MT034
c	16	61	67.8	349	12	BI021835	BI021835 RC5-MT025
c	17	61	67.8	353	10	BF893380	BF893380 QV3-MT012
c	18	61	67.8	354	9	AA321384	AA321384 EST23855
	19	61	67.8	354	9	AA321385	AA321385 EST23856
	20	61	67.8	355	10	BF893379	BF893379 QV3-MT012
	21	61	67.8	360	12	BI021822	BI021822 RC5-MT025
	22	61	67.8	370	10	BF893383	BF893383 QV3-MT012
	23	61	67.8	372	10	BF893378	BF893378 QV3-MT012
c	24	61	67.8	377	9	AA321224	AA321224 EST23726
c	25	61	67.8	379	14	BF999980	BF999980 CM3-MT019
c	26	61	67.8	379	14	C01361	C01361 HDMGS000834
c	27	61	67.8	387	12	BI021837	BI021837 RC5-MT025
c	28	61	67.8	388	10	BF896128	BF896128 CM2-MT015
c	29	61	67.8	389	12	BI021710	BI021710 CM3-MT034
c	30	61	67.8	396	12	BI021710	BI021710 CM3-MT034
c	31	61	67.8	413	10	AW468012	AW468012 he30d11.x
c	32	61	67.8	419	10	BF893435	BF893435 QV3-MT012
c	33	61	67.8	424	10	BF893232	BF893232 QV3-MT012
c	34	61	67.8	425	10	BF893231	BF893231 QV3-MT012
c	35	61	67.8	430	9	AI684407	AI684407 wa72d06.x
c	36	61	67.8	430	9	AI806882	AI806882 wf36d08.x
c	37	61	67.8	430	9	AA321410	AA321410 EST23882
c	38	61	67.8	433	10	BF893440	BF893440 QV3-MT012
c	39	61	67.8	449	10	BF893226	BF893226 QV3-MT012
c	40	61	67.8	450	10	BF893220	BF893220 QV3-MT012
c	41	61	67.8	465	9	AI312619	AI312619 qp77c06.x
c	42	61	67.8	472	10	BF893438	BF893438 QV3-MT012
c	43	61	67.8	473	10	BF900418	BF900418 RC4-MT023
c	44	61	67.8	475	10	BF893227	BF893227 QV3-MT012
c	45	61	67.8	486	9	AV741509	AV741509 QV3-MT012

ALIGNMENTS

RESULT 1
BF895944/c
BF895944

LOCUS   BF895944   221 bp   mRNA   linear   EST 18-JAN-2001
DEFINITION CM2-MT0157-221100-548-b02 MT0157 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF895944
VERSION   BF895944.1 GI:12287403
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 221)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

REFERENCE
  AUTHORS   Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   MEDLINE 20202663
   PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/getthtml2.pl?tl=CM2&t2=MT0157-
            22100-548-b02&t3=2000-11-22&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 54
            High quality sequence stop: 221.
FEATURES   Location/Qualifiers
     source   1..221
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0157"
                     /note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT   43 a   64 c   68 g   46 t
ORIGIN

Alignment Scores:
Pred. No.:   0.849
Score: 63.00
Percent Similarity: 84.21%
Best Local Similarity: 78.95%
Query Match: 70.00%

Length: 221
Matches: 15
Conservative: 1
Mismatches: 3
Indels: 0
Gaps: 0
DB: 10

US-10-045-180A-4 (1-19) x BF895944 (1-221)

```
Qy      1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpProAla 19
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    140 ATGAGGACCCTCGCCACTCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCT 84
```

RESULT 2
BI023130
LOCUS   BI023130   168 bp   mRNA   linear EST 14-JUN-2001
DEFINITION CM0-MT0390-070201-767-b09 MT0390 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI023130
VERSION   BI023130.1 GI:14429760
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 168)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

---

  TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   MEDLINE 20202663
   PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/getthtml2.pl?tl=CM0-MT0390-
            070201-767-b09&t3=2001-02-07&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 15
            High quality sequence stop: 168.
FEATURES   Location/Qualifiers
     source   1..168
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0390"
                     /note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT   30 a   59 c   43 g   36 t
ORIGIN

Alignment Scores:
Pred. No.:   1.21
Score: 61.00
Percent Similarity: 84.21%
Best Local Similarity: 73.68%
Query Match: 67.78%

Length: 168
Matches: 14
Conservative: 2
Mismatches: 3
Indels: 0
Gaps: 0
DB: 12

US-10-045-180A-4 (1-19) x BI023130 (1-168)

```
Qy      1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpProAla 19
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 ATGAGGACCCTCGCCACTCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCT 157
```

RESULT 3
BG193795
LOCUS   BG193795   194 bp   mRNA   linear EST 21-APR-2001
DEFINITION RST12931 Atherogs RAGE library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG193795
VERSION   BG193795.1 GI:13715482
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 194)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
            , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
            , E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
            ,J., Danzig,J. and Ducar,M.
  TITLE   Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
   MEDLINE 21227151

PUBMED   11329013
COMMENT
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 144.

FEATURES   Location/Qualifiers
     source   1..194
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."

BASE COUNT   38 a   63 c   56 g   37 t
ORIGIN

Alignment Scores:
Pred. No.:   1..49   Length: 194
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 67.78%   Indels: 0
DB: 10   Gaps: 0

Qy

US-10-045-180A-4 (1-19) x BG193795 (1-194)

Db

```
1  MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaThrTrpAla 19
   |||||||||||||||||||||||||||   ::|:||||    ::|:|||
18 ATGAGGACTCTGCGCCATCCTCGGCTGCCATTCTCCTCGGTGGCCCTGCAGGCCCAGGCT 74
```

RESULT 4
LOCUS   BQ334320
DEFINITION QV3-MT0129-111100-419-c11 MT0129 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ334320
VERSION   BQ334320.1 GI:20989401
KEYWORDS   EST.
SOURCE   Homo sapiens
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

BQ334320   243 bp   mRNA   linear   EST 20-MAY-2002

REFERENCE   1 (bases 1 to 243)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
            M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
PUBMED   10737800
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This sequence can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3-MT0129-
111100-419-c11&t3=2000-11-11&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 19.

FEATURES   Location/Qualifiers
     source   1..243
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MT0129"
                     /dev_stage="Adult"
                     /note="Organ: marrow; Vector: pUc18; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."

BASE COUNT   50 a   71 c   76 g   46 t
ORIGIN

Alignment Scores:
Pred. No.:   2..05   Length: 243
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 67.78%   Indels: 0
DB: 13   Gaps: 0

Qy

US-10-045-180A-4 (1-19) x BQ334320 (1-243)

Db

```
1   MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaThrTrpAla 19
    |||||||||||||||||||||||||||   ::|:||||    ::|:|||
156 ATGAGGACTCTGCGCCATCCTTGCTGCCATTCTCCTCGGTGGCCCTGCAGGCCCAGGCT 100
```

RESULT 5
LOCUS   AA321190
DEFINITION EST92992 Bone marrow Homo sapiens cDNA 5' end similar to defensin
AA321190   251 bp   mRNA   linear   EST 19-APR-1997
ACCESSION   AA321190
VERSION   AA321190.1 GI:1973837
KEYWORDS   EST.
SOURCE   Homo sapiens
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 251)
AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,
            C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,
            O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,
            L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-P., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,
            M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
TITLE   Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE   96026280
PUBMED   7566098
COMMENT
Other_ESTs: THC169174
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

```
FEATURES             Location/Qualifiers
     source          1..251
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):122061"
                     /sex="mixed"
                     /tissue_type="bone marrow"
                     /dev_stage="adult"
                     /clone_lib="Bone marrow"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       48 a     79 c     73 g     50 t     1 others
ORIGIN
```

US-10-045-180A-4 (1-19) x AA321190 (1-251)

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaTrpAla 19

Db   76 ATGAGGACCCTCGCCATCCTTGCTGCCCATTCTTCCTGGTGGCCCTGCAGGCCCAGGCT 132
```

RESULT 6

```
LOCUS       AA321281          281 bp    mRNA        linear    EST 19-APR-1997
DEFINITION  EST23811 Bone marrow Homo sapiens cDNA 5' end similar to defensin
            , mRNA sequence.
ACCESSION   AA321281
VERSION     AA321281.1  GI:1973668
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 281)
AUTHORS     Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseline,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
TITLE       Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
JOURNAL     Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE     96026280
PUBMED      7566098
COMMENT     Other_ESTs: THC169174
```

Alignment Scores:
```
Pred. No.:                2.15
Score:                   61.00
Percent Similarity:      84.21%     Length:            251
Best Local Similarity:   73.68%     Matches:            14
Query Match:             67.78%     Conservative:        2
DB:                          9      Mismatches:          3
                                    Indels:              0
                                    Gaps:                0
```

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

```
FEATURES             Location/Qualifiers
     source          1..281
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):121891"
                     /sex="mixed"
                     /tissue_type="bone marrow"
                     /dev_stage="adult"
                     /clone_lib="Bone marrow"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       60 a     87 c     78 g     56 t
ORIGIN
```

US-10-045-180A-4 (1-19) x AA321281 (1-281)

```
Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaTrpAla 19
        |||||||||||||||||||||||||:::||||:::||||||||||||||||||||||||||||
Db   68 ATGAGGACCCTCGCCATCCTTGCTGCCCATTCTTCCTGGTGGCCCTGCAGGCCCAGGCT 124
```

RESULT 7

```
LOCUS       AA321317          289 bp    mRNA        linear    EST 19-APR-1997
DEFINITION  EST23784 Bone marrow Homo sapiens cDNA 5' end similar to defensin
            , mRNA sequence.
ACCESSION   AA321317
VERSION     AA321317.1  GI:1973644
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 289)
AUTHORS     Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseline,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
TITLE       Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
JOURNAL     Nature 377 (6547 Suppl), 3-174 (1995)
```

Alignment Scores:
```
Pred. No.:                2.52
Score:                   61.00
Percent Similarity:      84.21%     Length:            281
Best Local Similarity:   73.68%     Matches:            14
Query Match:             67.78%     Conservative:        2
DB:                          9      Mismatches:          3
                                    Indels:              0
                                    Gaps:                0
```

Column 1

MEDLINE   96026280
PUBMED   7566098
COMMENT   Other_ESTs: THC169174
           Contact: Kerlavage, AR
           Bioinformatics
           The Institute for Genomic Research
           9712 Medical Center Drive, Rockville, MD 20850 USA
           Tel: 3018699056
           Fax: 3018699423
           Email: arkerlav@tigr.org
           For clone availability, additional sequence and expression
           information related to this EST, please check the TIGR Human Gene
           Index (http://www.tigr.org/tdb/hgi/hgi.html)
           Seq primer: M13 Reverse.

FEATURES   Location/Qualifiers
     source   1..289
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):121867"
                     /db_xref="taxon:9606"
                     /sex="mixed"
                     /tissue_type="bone marrow"
                     /dev_stage="adult"
                     /clone_lib="Bone marrow"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"

BASE COUNT   60 a   88 c   81 g   58 t   2 others
ORIGIN

Alignment Scores:
Pred. No.:   2.63
Score: 61.00
Percent Similarity: 84.21%
Best Local Similarity: 73.68%
Query Match: 67.78%

           Length: 289
           Matches: 14
           Conservative: 2
           Mismatches: 3
           Indels: 0
           Gaps: 0

US-10-045-180A-4 (1-19) x AA321317 (1-289)

Qy   1 MetArgThrLeuThrLeuuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
       ||||||||||| ::|||||||
Db 69 ATGAGGACCCTGACCCTGCCATTCCCTGGTGGCCCTGCACGCCCAGGCT 125

RESULT 8
BF99113/c
LOCUS   BF99113   290 bp   mRNA   linear
DEFINITION   QV3-GN0107-111100-423-a12 GN0107 Homo sapiens cDNA,
             mRNA sequence.
ACCESSION   BF99113
VERSION   BF99113.1 GI:12435228
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 290)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
  TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   MEDLINE   20202663
   PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil)
             Tel: +55-11-2704922
             Fax: +55-11-2707001

Column 2

           Email: asimpson@ludwig.org.br
           This sequence was derived from the FAPSP/LICR Human Cancer Genome
           Project. This entry can be seen in the following URL
           (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3&t2=QV3&t2=GN0107-
           11100-423-a12&t3=2000-11-11&t4=1)
           Seq primer: puc 18 forward
           High quality sequence stop: 284.

FEATURES   Location/Qualifiers
     source   1..290
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="GN0107"
                     /note="Organ: placenta normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 - Ludwig Institute for Cancer
                     Research) profiles into the pUC 18 vector; Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."

BASE COUNT   58 a   79 c   88 g   65 t
ORIGIN

Qy   1 MetArgThrLeuThrLeuuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
       ||||||||||| ::|||||||
Db 255 ATGAGGACCCTGACCCTGCCATTCCCTGGTGGCCCTGCACGCCCAGGCT 199

RESULT 9
AA321316
LOCUS   AA321316   294 bp   mRNA   linear
DEFINITION   EST23783 Bone marrow Homo sapiens cDNA 5' end similar to defensin
             1, mRNA sequence.
ACCESSION   AA321316
VERSION   AA321316.1 GI:1973643
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 294)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
             ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
             ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
             Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
             ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
             Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
             Kelley,J.C., Liu,L-I., Marmaros,S.M., Merrick,J.M.,
             Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
             Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
             Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
             Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
             Dimke,D., Feng,D-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
             , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
             Kusch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
             Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
             ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
             Venter,J.C.
  TITLE   Initial assessment of human gene diversity and expression patterns
             based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
   MEDLINE   96026280
   PUBMED   7566098

Alignment Scores:
Pred. No.:   2.64
Score: 61.00
Percent Similarity: 84.21%
Best Local Similarity: 73.68%
Query Match: 67.78%

           Length: 290
           Matches: 14
           Conservative: 2
           Mismatches: 3
           Indels: 0
           Gaps: 0

US-10-045-180A-4 (1-19) x BF99113 (1-290)

COMMENT   Other ESTs: THC169174
           Contact: Kerlavage, AR
           Bioinformatics
           The Institute for Genomic Research
           9712 Medical Center Drive, Rockville, MD 20850 USA
           Tel: 3018699056
           Fax: 3018699423
           Email: arkerlav@tigr.org
           For clone availability, additional sequence and expression
           information related to this EST, please check the TIGR Human Gene
           Index (http://www.tigr.org/tdb/hgi/hgi.html)
           Seq primer: M13 Reverse.

FEATURES   Location/Qualifiers
     source   1..294
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):121866"
                     /tissue_type="bone marrow"
                     /sex="mixed"
                     /dev_stage="adult"
                     /clone_lib="Bone marrow"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"

BASE COUNT   63 a   89 c   82 g   60 t
ORIGIN

US-10-045-180A-4 (1-19) x AA321316 (1-294)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        |||||||||||||| :||||||||||||||||||||||||||||||||
Db 61 ATGAGGACCCTCGCCATCCTTGCTGCCACCATTCTCCTGGTGGCCCTGCAGGCCCAGGCT 117

Alignment Scores:
Pred. No.:   2.69
Score: 61.00
Percent Similarity: 84.21%
Best Local Similarity: 73.68%
Query Match: 67.78%
DB: 9

Length: 294
Matches: 14
Conservative: 2
Mismatches: 3
Indels: 0
Gaps: 0

RESULT 10
LOCUS   AA321377   298 bp   mRNA   linear   EST 19-APR-1997
DEFINITION EST23847 Bone marrow Homo sapiens cDNA 5' end similar to defensin
            1, mRNA sequence.
ACCESSION   AA321377
VERSION   AA321377.1 GI:1973703
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 298)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE   Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence

---

JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE   96026280
PUBMED   7566098
COMMENT   Other ESTs: THC169174
          Contact: Kerlavage, AR
          Bioinformatics
          The Institute for Genomic Research
          9712 Medical Center Drive, Rockville, MD 20850 USA
          Tel: 3018699056
          Fax: 3018699423
          Email: arkerlav@tigr.org
          For clone availability, additional sequence and expression
          information related to this EST, please check the TIGR Human Gene
          Index (http://www.tigr.org/tdb/hgi/hgi.html)
          Seq primer: M13 Reverse.

FEATURES   Location/Qualifiers
     source   1..298
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):121926"
                     /tissue_type="bone marrow"
                     /sex="mixed"
                     /dev_stage="adult"
                     /clone_lib="Bone marrow"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"

BASE COUNT   63 a   90 c   83 g   61 t   1 others
ORIGIN

US-10-045-180A-4 (1-19) x AA321377 (1-298)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        |||||||||||||| :||||||||||||||||||||||||||||||||
Db 76 ATGAGGACCCTCGCCATCCTTGCTGCCACCATTCTCCTGGTGGCCCTGCAGGCCCAGGCT 132

Alignment Scores:
Pred. No.:   2.74
Score: 61.00
Percent Similarity: 84.21%
Best Local Similarity: 73.68%
Query Match: 67.78%
DB: 9

Length: 298
Matches: 14
Conservative: 2
Mismatches: 3
Indels: 0
Gaps: 0

RESULT 11
LOCUS   AA321102   312 bp   mRNA   linear   EST 19-APR-1997
DEFINITION EST23896 Bone marrow Homo sapiens cDNA 5' end similar to defensin
            1, mRNA sequence.
ACCESSION   AA321102
VERSION   AA321102.1 GI:1973750
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 312)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and

First record

```
TITLE       Venter,J.C.
            Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
JOURNAL     Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE     96026280
PUBMED      7566098
COMMENT     Other ESTs: THC169174
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..312
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):121974"
                     /db_xref="taxon:9606"
                     /sex="mixed"
                     /tissue_type="bone marrow"
                     /dev_stage="adult"
                     /clone_lib="Bone marrow"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      66 a     94 c     90 g     61 t      1 others
ORIGIN

US-10-045-180A-4 (1-19) x AA321102 (1-312)

Qy          1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaAla 19
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 ATGAGGACCCTCGCCATCCTTGCTGCTGGTGGCGCTGCAGGCCCAGGCT 135
```

```
RESULT 12
LOCUS       AA321426                 320 bp    mRNA    linear    EST 19-APR-1997
DEFINITION  EST23923 Bone marrow Homo sapiens cDNA 5' end similar to defensin
            1, mRNA sequence.
ACCESSION   AA321426
VERSION     AA321426.1  GI:1973774
KEYWORDS    EST.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 320)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
```

Alignment Scores (RESULT 12):

```
Alignment Scores:
Pred. No.:
Score:                   2.93        Length:        312
Percent Similarity:     61.00        Matches:        14
Best Local Similarity:  84.21%       Conservative:    2
Query Match:            73.68%       Mismatches:      3
                        67.78%       Indels:          0
                         9           Gaps:            0
```

Second record

```
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
TITLE       Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
JOURNAL     Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE     96026280
PUBMED      7566098
COMMENT     Other ESTs: THC169174
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..320
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):121998"
                     /db_xref="taxon:9606"
                     /sex="mixed"
                     /tissue_type="bone marrow"
                     /dev_stage="adult"
                     /clone_lib="Bone marrow"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      69 a     93 c     88 g     65 t      5 others
ORIGIN

US-10-045-180A-4 (1-19) x AA321426 (1-320)

Qy          1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAlaAla 19
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 ATGAGGACCCTCGCCATCCTTGCTGCTGGTGGCGCTGCAGGCCCAGGCT 126
```

```
RESULT 13
LOCUS       AA321192                 327 bp    mRNA    linear    EST 19-APR-1997
DEFINITION  EST23994 Bone marrow Homo sapiens cDNA 5' end similar to defensin
            1, mRNA sequence.
ACCESSION   AA321192
VERSION     AA321192.1  GI:1973839
KEYWORDS    EST.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 327)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
```

Alignment Scores (RESULT 13):

```
Alignment Scores:
Pred. No.:
Score:                   3.04        Length:        320
Percent Similarity:     61.00        Matches:        14
Best Local Similarity:  84.21%       Conservative:    2
Query Match:            73.68%       Mismatches:      3
                        67.78%       Indels:          0
                         9           Gaps:            0
```

TITLE　　　　　Shotgun sequencing of the human transcriptome with ORF expressed
　　　　　　　　sequence tags

REFERENCE
AUTHORS　　Simpson,A.J.
　　　　　　　　,M.J.,Soares,F.,Brentani,R.R.,Reis,L.F.,de Souza,S.J. and
　　　　　　　　Brunstein,A.,deOliveira,P.S.,Bucher,P.,Jongeneel,C.V.,O'Hare
　　　　　　　　Goldman,G.H.,Carvalho,A.F.,Matsukuma,A.,Baia,G.S.,Simpson,F.F.,
　　　　　　　　Nagai,M.A.,da Silva,W. Jr.,Zago,M.A.,Bordin,S.,Costa,F.F.,
　　　　　　　　Dias Neto,E.,Garcia Correa,R.,Verjovski-Almeida,S.,Briones,M.R.,
　　　　　　　　1　(bases 1 to 348)

Qy　　　　　1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
　　　　　　　　|||||||||||||||||||||||||||||||||||
Db　　　　　75 ATGAGGACCCTCGCCATCCTTGCTGCCATTCTCCTGGTGGCGCTGCAGGCCCAGGCT 131

US-10-045-180A-4 (1-19) x AA321192 (1-327)

Alignment Scores:
Pred. No.:　　　　　　　3.13
Score:　　　　　　　　　61.00
Percent Similarity:　　84.21%
Best Local Similarity:　73.68%
Query Match:　　　　　9

DB:

BASE COUNT　　72 a　97 c　91 g　66 t　1 others
ORIGIN

FEATURES　　　Location/Qualifiers
source　　　　　1..327
　　　　　　　　/organism="Homo sapiens"
　　　　　　　　/mol_type="mRNA"
　　　　　　　　/db_xref="ATCC (inhost):122063"
　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　/sex="mixed"
　　　　　　　　/tissue_type="bone marrow"
　　　　　　　　/dev_stage="adult"
　　　　　　　　/clone_lib="Bone marrow"
　　　　　　　　/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
　　　　　　　　XhoI"

Qy　　　　　1..327

RESULT 14
LOCUS　　　　BF893385　　　　348 bp　　　mRNA　　　linear　　　EST 18-JAN-2001
DEFINITION　QV3-MT0129-111100-427-h01 MT0129 Homo sapiens cDNA, mRNA sequence.
ACCESSION　　BF893385
VERSION　　　BF893385.1　GI:12284844
KEYWORDS　　EST.
SOURCE　　　Homo sapiens (human)
ORGANISM　　Homo sapiens
　　　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　　1　(bases 1 to 348)
AUTHORS　　Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
　　　　　　　　Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
　　　　　　　　Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
　　　　　　　　Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
　　　　　　　　,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
　　　　　　　　Simpson,A.J.
TITLE　　　　　Shotgun sequencing of the human transcriptome with ORF expressed
　　　　　　　　sequence tags

JOURNAL　　　Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE　　　20202663
PUBMED　　　10737800
COMMENT　　　Contact: Simpson A.J.G.
　　　　　　　　Laboratory of Cancer Genetics

REFERENCE
AUTHORS　　Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
　　　　　　　　Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
　　　　　　　　, Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
　　　　　　　　Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
　　　　　　　　Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
　　　　　　　　,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
　　　　　　　　Venter,J.C.
TITLE　　　　　Initial assessment of human gene diversity and expression patterns
　　　　　　　　based upon 83 million nucleotides of cDNA sequence
JOURNAL　　　Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE　　　96026280
PUBMED　　　7566098
COMMENT　　　Other ESTs: THC169174
　　　　　　　　Contact: Kerlavage, AR
　　　　　　　　Bioinformatics
　　　　　　　　The Institute for Genomic Research
　　　　　　　　9712 Medical Center Drive, Rockville, MD 20850 USA
　　　　　　　　Tel: 301899056
　　　　　　　　Fax: 301869423
　　　　　　　　Email: arkerlav@tigr.org
　　　　　　　　For clone availability, additional sequence and expression
　　　　　　　　information related to this EST, please check the TIGR Human Gene
　　　　　　　　Index [http://www.tigr.org/tdb/hgi/hgi.html]
　　　　　　　　Seq primer: M13 Reverse.

FEATURES　　　Location/Qualifiers
source　　　　　1..327
　　　　　　　　/organism="Homo sapiens"
　　　　　　　　/mol_type="mRNA"
　　　　　　　　/db_xref="ATCC (inhost):122063"
　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　/sex="mixed"
　　　　　　　　/tissue_type="bone marrow"
　　　　　　　　/dev_stage="adult"
　　　　　　　　/clone_lib="Bone marrow"
　　　　　　　　/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
　　　　　　　　XhoI"

COMMENT　　　Laboratory of Cancer Genetics

JOURNAL　　　Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE　　　20202663
PUBMED　　　10737800
TITLE　　　　　Shotgun sequencing of the human transcriptome with ORF expressed
　　　　　　　　sequence tags
　　　　　　　　Simpson,A.J.
　　　　　　　　,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
　　　　　　　　Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
　　　　　　　　Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
　　　　　　　　Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
REFERENCE　　Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
AUTHORS　　　1　(bases 1 to 349)
　　　　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
　　　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM　　Homo sapiens
SOURCE　　　Homo sapiens (human)
KEYWORDS　　EST.
VERSION　　　BI021617.1　GI:14428247
ACCESSION　　BI021617
DEFINITION　CM3-MT0343-170101-665-a05 MT0343 Homo sapiens cDNA, mRNA sequence.
LOCUS　　　　BI021617/c
BI021617/c　BI021617　　　　349 bp　　　mRNA　　　linear　　　EST 14-JUN-2001
RESULT 15

Db　　　　　87 ATGAGGACCCTCGCCATCCTTGCTGCCATTCTCCTGGTGGCGCTGCAGGCCCAGGCT 143
　　　　　　　　|||||||||||　 :::||||||||||||||||||||||||||||||||||| |||
Qy　　　　　1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19

US-10-045-180A-4 (1-19) x BF893385 (1-348)

DB:　　　　　　　　　　10
Best Local Similarity:　73.68%　　Gaps:　　　　0
Percent Similarity:　　84.21%　　Mismatches:　3
Score:　　　　　　　　　61.00　　Conservative:　2
Pred. No.:　　　　　　　3.42　　Matches:　　　14
Alignment Scores:　　　　　　　　Length:　　　348

ORIGIN
BASE COUNT　　79 a　103 c　97 g　69 t

FEATURES　　　Location/Qualifiers
source　　　　　1..348
　　　　　　　　/organism="Homo sapiens"
　　　　　　　　/mol_type="mRNA"
　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　/dev_stage="Adult"
　　　　　　　　/clone_lib="MT0129"
　　　　　　　　/note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
　　　　　　　　SmaI; A mini-library was made by cloning products derived
　　　　　　　　from ORESTES PCR (U.S. Letters Patent application No. 196
　　　　　　　　,716 - Ludwig Institute for Cancer Research) profiles
　　　　　　　　into the pUC 18 vector. Reverse transcription of tissue
　　　　　　　　mRNA and cDNA amplification were performed under low
　　　　　　　　stringency conditions."
High quality sequence stop: 348.
　　　　　　　　Seq primer: puc 18 forward
111100-427-h01&t3=2000-11-11&t4=1]
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3-MT0129-
Project. This entry can be seen in the following URL
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Email: asimpson@ludwig.org.br
Fax: +55-11-2707001
Tel: +55-11-2704922
Brazil
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Ludwig Institute for Cancer Research
Laboratory of Cancer Genetics
Contact: Simpson A.J.G.
COMMENT
PUBMED　　　10737800
MEDLINE　　　20202663
JOURNAL　　　Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-27049922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM3&t2=CM3-MT0343-
170101-665-a05&t3=2001-01-17&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 349.

FEATURES   Location/Qualifiers
    source   1..349
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /dev_stage="Adult"
                        /clone_lib="MT0343"
                        /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                        SmaI; A mini-library was made by cloning products derived
                        from   ORESTES PCR (U.S. Letters Patent application No. 196
                        ,716 - Ludwig Institute for Cancer Research) profiles
                        into   the pUC 18 vector. Reverse transcription of tissue
                        mRNA and cDNA amplification were performed under low
                        stringency conditions."

BASE COUNT   69 a   98 c   98 g   84 t
ORIGIN

US-10-045-180A-4 (1-19) x BI021617 (1-349)

DB:
Query Match: 12   Gaps: 0
Best Local Similarity: 67.78%   Indels: 0
Percent Similarity: 73.68%   Mismatches: 3
Score: 84.21%   Conservative: 2
Pred. No.:   61.00   Matches: 14
                        3.43   Length: 349

Alignment Scores:

Qy   1 MetArgThrIleuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
           |||||||||||||||| ::::||:|| :::||||||||||||||||||| |||
Db   311 ATGAGGACCCTGCCAACCTTGCCATCCTTCCTGGTGGCCCTGCAGGGCCCAGGCT 255

Search completed: December 17, 2003, 18:25:43
Job time : 665.293 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: December 17, 2003, 14:50:33 ; Search time 71.25 Seconds
(without alignments)
719.850 Million cell updates/sec

Title: US-10-045-180A-4
Perfect score: 90
Sequence: 1 MRTLTLLSAFLIVALQAWA 19

Scoring table: BLOSUM62
Xgapop 10.0 : Xgapext 0.5
Ygapop 10.0 : Ygapext 0.5
Fgapop 6.0 : Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145023_12520/app_query.fasta_1.860
-DB=N Geneseq 19Jun03 -OFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER=US10045180 @CGN_1_1_835 @runat_17122003_145023_12520 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
N_Geneseq 19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	90	100.0	453	20	AAX26697	cDNA sequence enco
2	90	100.0	453	20	AAX26698	cDNA sequence enco
3	90	100.0	4415	20	AAX26696	Nucleotide sequence
4	64	71.1	294	22	AAH57220	Human intestine sp
5	64	71.1	398	24	ABN59997	Novel human coding
6	64	71.1	452	14	AAQ53217	Sequence encoding
7	64	71.1	452	25	ACA56600	Human signalling p
8	61	67.8	110	17	AAT33496	110mer used in con
9	61	67.8	110	18	AAT63059	110mer used in co
10	61	67.8	110	19	AAV10895	Human pre-pro cart
11	61	67.8	187	22	ABA75757	Human foetal liver
12	61	67.8	187	22	AAK50398	Human bone marrow
13	61	67.8	187	24	ABS23860	Human genome-deriv
14	61	67.8	299	16	AAT26106	Human gene signatu
15	61	67.8	464	21	AAF20911	Human defensin 3 p
16	61	67.8	464	21	AAA34789	Human defensin 3 p
17	61	67.8	478	21	AAC04468	Human adenosine re
18	61	67.8	478	12	AAQ14908	Chronic myelogenou
19	61	67.8	498	18	ABQ80601	Human defensin 1 c
20	61	67.8	514	24	ABN59722	Novel human coding
21	61	67.8	597	22	ABA63315	Human foetal liver
22	61	67.8	597	22	AAK37526	Human bone marrow
23	61	67.8	597	24	ABS11502	Human genome-deriv
24	61	67.8	644	21	AAF21446	Human defensin 3 p
25	61	67.8	652	24	AB211202	Human polynucleoti
26	61	67.8	664	24	AB211183	Human polynucleoti
27	61	67.8	3710	21	AAF20910	Human defensin 3 p
28	61	67.8	3710	21	AAA34788	Human defensin 3 p
29	61	67.8	3710	24	ABQ80602	Human defensin 1 r
30	61	67.8	3710	21	ABK83829	Human cDNA differe
31	61	67.8	8365	21	AAF21445	Human defensin 2 p
32	61	67.8	209273	21	AAF21437	Human factor-relat
33	c 60	66.7	473	25	ABZ36740	Human GENSET codin
34	c 59	65.6	411	24	ABQ60009	Human colon cancer
35	c 59	65.6	424	14	ABQ60783	Sequence encoding
36	59	65.6	449	24	AAQ53216	Human defensin 5 D
37	59	65.6	543	25	ABZ36767	Human GENSET codin
38	c 59	65.6	664	15	AAO68911	Human c-myc far up
39	c 59	65.6	1097	17	AAQ53218	Genomic sequence e
40	c 58	64.4	92	18	AAT33499	92mer used in con
41	c 58	64.4	92	19	AAT63062	92-mer used in con
42	c 58	64.4	92	19	AAV10998	92mer used in cart
43	c 56	62.2	105	19	AAV60903	Human pre-pro cart
44	c 55	61.1	537	24	ABN65503	DNA fragment SYN-U
45	c 53	58.9	480	25	AAD47375	Human LP321 DNA.

ALIGNMENTS

RESULT 1
AAX26697

ID   AAX26697 standard; cDNA; 453 BP.
XX
AC   AAX26697;
XX
DT   18-JUN-1999   (first entry)
XX
DE   cDNA sequence encoding human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis; ss.
XX
OS   Homo sapiens.

PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;   97FR-0010823.
XX
PR   29-AUG-1997;   97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
XX
DR   WPI; 1999-183266/16.
DR   P-PSDB; AAY01604.
XX
PS   Disclosure; Fig 3; 56pp; French.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
XX
CC   The present sequence encodes human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 453 BP; 107 A; 121 C; 97 G; 128 T; 0 other;

US-10-045-180A-4 (1-19) x AAX26697 (1-453)

Alignment Scores:
Pred. No.:   1.09e-06
Score: 90.00
Percent Similarity: 100.00%   Length: 453
Best Local Similarity: 100.00%   Matches: 19
Query Match: 20   Conservative: 0
DB: Mismatches: 0
                                           Indels: 0
                                           Gaps: 0

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   52 ATGAGGACCCTCACCCTCCTCTCGGCCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCA 108

RESULT 2
AAX26698
ID   AAX26698 standard; cDNA; 453 BP.
XX
AC   AAX26698;
XX
DT   18-JUN-1999   (first entry)
XX
DE   cDNA sequence encoding human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis; ss.
XX
OS   Homo sapiens.
XX

PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;   97FR-0010823.
XX
PR   29-AUG-1997;   97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
XX
DR   WPI; 1999-183266/16.
XX
PS   Claim 5; Fig 4; 56pp; French.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
XX
CC   The present sequence encodes human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 453 BP; 107 A; 121 C; 97 G; 128 T; 0 other;

US-10-045-180A-4 (1-19) x AAX26698 (1-453)

Alignment Scores:
Pred. No.:   1.09e-06
Score: 90.00
Percent Similarity: 100.00%   Length: 453
Best Local Similarity: 100.00%   Matches: 19
Query Match: 20   Conservative: 0
DB: Mismatches: 0
                                           Indels: 0
                                           Gaps: 0

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   52 ATGAGGACCCTCACCCTCCTCTCGGCCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCA 108

RESULT 3
AAX26696
ID   AAX26696 standard; DNA; 4415 BP.
XX
AC   AAX26696;
XX
DT   18-JUN-1999   (first entry)
XX
DE   Nucleotide sequence of human defensin (Def-X).
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.

```
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;  97FR-0010823.
XX
PR   29-AUG-1997;  97FR-0010823.
XX
PA   (GEST ) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Claim 5; Fig 2; 56pp; French.
XX
CC   The present sequence encodes human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 4415 BP; 1128 A; 1109 C; 912 G; 1252 T; 14 other;

Qy   US-10-045-180A-4 (1-19) x AAX26696 (1-4415)
DB:

Alignment Scores:
Pred. No.:                1.79e-05          Length:        4415
Score:                    90.00             Matches:       19
Percent Similarity:       100.00%           Conservative:  0
Best Local Similarity:    100.00%           Mismatches:    0
Query Match:              100.00%           Indels:        0
                          20                Gaps:          0

Qy      1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3406 ATGAGGACCCTCACCCTCCTCTCTGCCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCA 3462
```

```
RESULT 4
AAH57220
ID   AAH57220 standard; cDNA; 294 BP.
XX
AC   AAH57220;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human intestine specific cDNA sequence SEQ ID NO:60.
XX
KW   Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
KW   lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW   metabolic disease; developmental disease; cytostatic; immunoregulatory;
KW   neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX
OS   Homo sapiens.
XX
PN   WO200132927-A2.
XX
PD   10-MAY-2001.
XX
PF   02-NOV-2000;  2000WO-US30396.
XX
PR   04-NOV-1999;  99US-0163508.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Sornasse T, Seilhamer JJ, Watson GA;
XX
DR   WPI; 2001-291057/30.
XX
PT   New cell and tissue specific polynucleotides useful for diagnosis,
PT   prognosis or monitoring of treatments for disorders where the gene is
PT   associated with a cancer, immunopathology or neuropathology -
XX
PS   Claim 1; Page 95; 327pp; English.
XX
CC   AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
CC   sequences (I). (I) can have cytostatic, immunomodulatory and
CC   neuroprotective activities, and can be used in gene therapy. (I) and
CC   proteins (II) encoded by then are used in high throughput screening
CC   assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC   mimetics, peptides, proteins, agonists, antagonists, antibodies or
CC   their fragments, immunoglobulins, inhibitors, drug compounds and
CC   pharmaceutical agents. Expression of (I) in a sample indicates the
CC   differentiation of embryonic stem cells into a tissue selected from
CC   brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC   tissue. (I) and (II) are used to produce an expression profile that
CC   defines a metabolic or developmental process, treatment, condition,
CC   disease or disorder. The gene profile can be used for diagnosis,
CC   prognosis or monitoring of treatments and for investigating a
CC   predisposition to a disorder where the gene is associated with a
CC   cancer, immunopathology or neuropathology.
XX
SQ   Sequence 294 BP; 72 A; 80 C; 73 G; 64 T; 5 other;

Qy   US-10-045-180A-4 (1-19) x AAH57220 (1-294)
DB:

Alignment Scores:
Pred. No.:                0.0247            Length:        294
Score:                    64.00             Matches:       15
Percent Similarity:       89.47%            Conservative:  2
Best Local Similarity:    78.95%            Mismatches:    2
Query Match:              71.11%            Indels:        0
                          22                Gaps:          0

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        ||||||||||||||:::||||:|:||::|||:||||||||||||||||||::||||
Db   15 ATGAGGACCCTCACCATCCTCACTGCTGTTCTCCTGCGTGGCCCTTCAGGCCAAGGCT 71
```

```
RESULT 5
ABN59997
ID   ABN59997 standard; cDNA; 398 BP.
XX
AC   ABN59997;
XX
DT   28-JUN-2002  (first entry)
XX
DE   Novel human coding sequence SEQ ID NO: 408.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; immunomodulator;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US26015.
```

XX
XX
PR
PA   11-SEP-2000; 200US-0659671.
XX
XX   (HYSE-) HYSEQ INC.
PA
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-292408/33.
PI
DR   P-PSDB; ABB97584.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Claim 1; SEQ ID NO 408; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat nervous system disorders
CC   e.g. rheumatoid arthritis, and to treat inflammatory conditions e.g.
CC   Parkinson's disease. The present sequence is a coding sequence of the
CC   invention.
XX
SQ   Sequence 398 BP; 102 A; 108 C; 93 G; 95 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 0.0359
Percent Similarity: 64.00   Length: 398
Percent Identity: 89.47%   Matches: 15
Best Local Similarity: 78.95%   Conservative: 2
Query Match: 71.11%   Mismatches: 2
                          24   Indels: 0
                                         Gaps: 0

US-10-045-180A-4 (1-19) x ABN59997 (1-398)

Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
         |||||||||||||||||||::|||::|||||||||||||||||||| |||
Db 30 ATGAGAACCCTCACCATTCCTCCACTCGTGTTCCTCGTGGCCCTCCAGGCCAAGGCT 86

US-10-045-180a-4

RESULT 6
ID   AAQ53217
AC   AAQ53217;
DT   25-MAR-2003 (updated)
DT   17-JUN-1994 (first entry)
DE   Sequence encoding gastrointestinal defensin (GID) peptide called
DE   human defensin 6.
KW   Gastrointestinal defensin peptide; GID; pharmaceutical; Paneth
KW   cell; antimicrobial; anti-inflammatory; diagnosis;
KW   contact disinfectant; ss.
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   19..321
FT   /*tag= a
XX
PN   WO9324513-A1.
XX
PD   09-DEC-1993.
PF   18-MAY-1993; 93WO-US047740.
XX
PR   22-MAY-1992; 92US-0888232.

XX
PA
XX   (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.
PI   Bevins CL, Jones DE;
XX
DR   WPI; 1993-405719/50.
DR   P-PSDB; AAR44819.
XX
PT   Gastrointestinal defensin peptide(s) - useful as antimicrobial
PT   and anti-inflammatory agents and for detecting gastrointestinal
PT   disorders
XX
PS   Claim 4; page 48-49; 97pp; English.
XX
CC   A probe (D5' oligo) having the sequence given in AAQ53219 (claimed)
CC   was used to screen a population of clones to identify novel defensin
CC   defensin peptides. Pref. the libraries were human genomic and cDNA
CC   libraries. Hybridisation and partial sequence analysis of the
CC   identified clones contained previously characterised myeloid derived
CC   defensin sequences as well as new defensin related sequences. Two
CC   clones expressing new defensin related sequences were extensively
CC   characterised and found to contain genes selectively expressed in
CC   Paneth cells of the small intestine. These Paneth cell-derived
CC   defensins are designated human defensin 5 and human defensin 6 and
CC   are referred to as gastrointestinal defensin peptides. Comparison
CC   of the deduced AA sequence of the defensin 5 cDNA with the
CC   previously reported preprodefensin shows significant similarity.
CC   The deduced AA sequence of defensin 6 cDNA has features similar to
CC   previously reported preprodefensins.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 452 BP; 123 A; 118 C; 98 G; 113 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 0.0419
Percent Similarity: 64.00   Length: 452
Percent Identity: 89.47%   Matches: 15
Best Local Similarity: 78.95%   Conservative: 2
Query Match: 71.11%   Mismatches: 2
                          14   Indels: 0
                                         Gaps: 0

US-10-045-180A-4 (1-19) x AAQ53217 (1-452)

Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
         |||||||||||||||||||::|||::|||||||||||||||||||| |||
Db 19 ATGAGAACCCTCACCATTCCTCCACTCGTGTTCCTCGTGGCCCTCCAGGCCAAGGCT 75

RESULT 7
ACA56600
ID   ACA56600 standard; cDNA; 452 BP.
AC   ACA56600;
XX
DT   06-JUN-2003 (first entry)
XX
DE   Human signalling pathway polynucleotide probe SBQ ID NO 1198.
KW   Human; probe; ss; array element; Parkinson's disease;
KW   signalling pathway population; cancer; adenocarcinoma; leukaemia;
KW   immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
OS   Homo sapiens.
XX
PN   US6500938-B1.
XX
PD   31-DEC-2002.
PF   30-JAN-1998; 98US-0016434.
PR   30-JAN-1998; 98US-0016434.
PA   (INCY-) INCYTE GENOMICS INC.
XX

PI   Au-Young J, Seilhamer JJ;
XX
DR   WPI; 2003-352189/33.
XX
PT   microarray for monitoring the expression of a number of target
PT   polynucleotides -
PT   polynucleotides
XX
PS   Claim 1; SEQ ID NO 1198; 65pp; English.
XX
CC   The invention relates to a combination, comprises a sequence selected from one of the 1490
CC   polynucleotide probes comprising a sequence selected from one of the 1490
CC   sequences mentioned in the specification. The combination is useful as an
CC   array element in a microarray for monitoring the expression of a number
CC   of target polynucleotides. The microarray is particularly useful in the
CC   diagnosis and treatment of cancer and immunopathology and neuropathology.
CC   The microarray is useful in diagnostics and treatment regimens, drug
CC   discovery and development, toxicological and carcinogenicity studies,
CC   forensics and pharmacogenomics. The microarray is also useful for
CC   monitoring progression of diseases and for developing sophisticated
CC   profiles for the effects of currently available therapeutic drugs. The
CC   combination is also useful for purifying a subpopulation of mRNAs, cDNAs
CC   and genomic fragments and in research and diagnostic applications. The
CC   array can detect changes in expression in a large number of genes coding
CC   for different signaling pathway populations which can be used to diagnose
CC   various diseases including cancer e.g. adenocarcinoma and leukaemia,
CC   immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
CC   and Parkinson's disease. The present sequence represents a polynucleotide
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=0650093B1.
XX
SQ   Sequence 452 BP; 124 A; 118 C; 98 G; 112 T; 0 other;

US-10-045-180A-4 (1-19) x ACA56600 (1-452)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          |||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db   19 ATGAGAACCCTCACCATCCTACTCGTGTTTCCTCGTGGCCCTCCAGGCCAAGGCT 75

Alignment Scores:
Pred. No.:   0.0419
Score: 64.00   Length: 452
Percent Similarity: 89.47%   Matches: 15
Best Local Similarity: 78.95%   Conservative: 2
Query Match: 71.11%   Mismatches: 2
DB: 25   Indels: 0
                                                       Gaps: 0

RESULT 8
ID   AAT33496
AC   AAT33496;
XX
DT   21-MAY-1997 (first entry)
XX
DE   110mer used in construction of pre pro cartridge.
XX
KW   Cationic peptide; CEMA, CEME; cecropia; melittin; bacterial growth;
KW   antibiotic activity; permeable; outer membrane;
KW   lipopolysaccharide; fusion peptide; inhibition; endotoxaemia;
KW   sepsis associated disorder; septic shock; ss.
XX
OS   Synthetic.
XX
PN   WO9628559-A1.
XX
PD   19-SEP-1996.
XX
PF   13-MAR-1996;   96WO-IB00431.
XX
PR   13-MAR-1995;   95US-0405234.

XX
PA   (UIBR-) UNIV BRITISH COLUMBIA.
XX
PI   Brown MH, Hancock REW, Kelly N, Piers KL;
XX
DR   WPI; 1996-433831/43.
XX
PT   Recombinant production of antimicrobial, cationic peptide - produced
PT   in bacterial host as fusion peptide with an anionic portion for
PT   suppressing the anti-microbial activity of the cationic peptide
XX
PS   Example 2; Page 19; 90pp; English.
XX
CC   The sequences given in AAT33488-500 and AAT40195-204 are
CC   oligonucleotides which were used in the construction of vectors used in
CC   the method of the invention. The method comprises constructing an
CC   expression vector containing DNA encoding the cationic peptide adjacent
CC   to DNA encoding an anionic carrier peptide which inhibits bacterial
CC   proteases. A bacterial host cell is transformed with the vector and
CC   cultured to express a fusion peptide comprising the cationic and anionic
CC   peptides. The biologically active cationic peptide is then recovered.
CC   The vectors constructed using these sequences encode CEME, CEMA or HNP-1
CC   (human neutrophil peptide 1). CEMA contains the first 18 amino acids of
CC   cecropin A and the last 8 residues of melittin. CEMA is derived from
CC   CEME by changing the C-terminal amino acid of CEME to include two
CC   additional Lys residues. This modification improves the antibiotic
CC   activity of the peptide against many bacterial species, two fold, as
CC   well as substantially enhancing its ability to permeabilise bacterial
CC   outer membranes and to bind to lipopolysaccharide. CEMA and CEME act
CC   to permeabilise bacterial membranes breaking down the outer membrane
CC   permeability barrier. They may be used in conjunction with antibiotics
CC   to inhibit bacterial growth and to treat an endotoxaemia or sepsis
CC   associated disorders e.g. septic shock.
XX
SQ   Sequence 110 BP; 21 A; 38 C; 32 G; 19 T; 0 other;

US-10-045-180A-4 (1-19) x AAT33496 (1-110)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          |||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db   5 ATGAGGAGCCCTCGCCATCCTTGCCTGCCCAATTCTCCTTGGTGGCCCTGCAGGCCCAGGCT 61

Alignment Scores:
Pred. No.:   0.025
Score: 61.00   Length: 110
Percent Similarity: 84.21%   Matches: 14
Best Local Similarity: 73.68%   Conservative: 2
Query Match: 67.78%   Mismatches: 3
DB: 17   Indels: 0
                                                       Gaps: 0

RESULT 9
ID   AAT63059
AC   AAT63059;
XX
DT   25-MAR-2003 (updated)
DT   08-MAY-1997 (first entry)
XX
DE   110-mer used in construction of pre pro cartridge.
XX
KW   HNP-1; CEME; CEMA; human neutrophil protein; fusion; insect defensin;
KW   cecropin A; bee venom; melittin; antibiotic activity; S. typhimurium;
KW   permeability; bacterial outer membrane; lipopolysaccharide;
KW   antibiotic; E. coli; Pseudomonas aeruginosa; E. cloacae; S. aureus; ss.
XX
OS   Synthetic.
XX
PN   US5593866-A.
XX
PD   14-JAN-1997.
XX
PF   20-DEC-1995;   95US-0575052.

XX
PR   20-AUG-1993;   93US-0110502.
PR   21-AUG-1992;   92US-0933492.
PR   20-DEC-1995;   95US-0575052.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI   Brown MH, Hancock REW, Piers KL;
XX
DR   WPI; 1997-099468/09.
XX
PT   Prodn. of recombinant cationic peptide with antimicrobial activity -
PT   in the form of a fusion peptide with protease-inhibiting anionic
PT   peptide
XX
PS   Example 1; Columnn 9-10; 31pp; English.
XX
CC   The sequences given in AAT63051-64 represent oligonucleotides which were
CC   used in the construction of pGST-pptide plasmids for the expression of
CC   HNP-1, CEME and CEMA. HNP-1 is peptide 1 from human neutrophil protein,
CC   CEME is a fusion peptide made from portions of an insect defensin
CC   cecropin A and the bee venom peptide melittin, and CEMA is a variant
CC   of CEME with two additional lysine residues at the C-terminal end.
CC   The modification from CEME to CEMA causes a two-fold improvement in
CC   antibiotic activity against many bacterial species as well as
CC   substantially enhancing the ability to permeabilise bacterial outer
CC   membranes and binding to lipopolysaccharides. CEMA may be used
CC   synergistically with classical antibiotics to break down the outer
CC   membrane permeability barrier. It is possibly active against E. coli,
CC   Pseudomonas aeruginosa, E. cloacae, S. typhimurium and S. aureus.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 110 BP; 21 A; 38 C; 32 G; 19 T; 0 other;

Alignment Scores:
Pred. No.:   0.025   Length: 110
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 18   Indels: 0
DB: Gaps: 0

US-10-045-180A-4 (1-19) x AAT63059 (1-110)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLaLeuGlnAlaAlaTrpAlaAla 19
          |||||||||||| ::||::||::||| |||||||||||||||||||
Db   5 ATGAGGACCCTGCGCCATCCTGCCATTCCTCGTGCTGGCCCTGCAGGCCCAGGCT 61

RESULT 10
AAV10895
ID   AAV10895 standard; DNA; 110 BP.
XX
AC   AAV10895;
XX
DT   14-JUL-1998 (first entry)
XX
DE   Human pre-pro cartridge primer 1.
XX
KW   Cecropin A; mellitin; CEME; fusion protein; polycationic; antibacterial;
KW   permeable; outer membrane; lipopolysaccharide; bacterial infection;
KW   septic shock; primer; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   US5707855-A.
XX
PD   13-JAN-1998.
XX
PF   20-DEC-1996;   96US-0770557.
XX
PR   20-AUG-1993;   93US-0110502.
PR   21-AUG-1992;   92US-0933492.

PR   20-DEC-1995;   95US-0575052.
PR   20-DEC-1996;   96US-0770557.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI   Brown MH, Hancock REW, Piers KL;
XX
DR   WPI; 1998-100348/09.
XX
PT   Antibacterial peptide CEMA - useful for treating bacterial
PT   infections and septic shock
XX
PS   Example 1; Column 10; 31pp; English.
XX
CC   AAV10887-V10909 are primers used in a method which allows the production
CC   of a novel polycationic antibacterial peptide, CEMA. This peptide is a
CC   is a hybrid comprising the first 18 amino acids of cecropin A and the
CC   last 8 amino acids of melittin. It has two Lys residues at the C
CC   terminus which creates a sequence with two-fold increased antibiotic
CC   activity against many bacterial species as well as enhanced ability to
CC   permeabilise bacterial outer membranes and bind to lipopolysaccharide.
CC   CEMA can be used to treat bacterial infections and septic shock.
XX
SQ   Sequence 110 BP; 21 A; 38 C; 32 G; 19 T; 0 other;

Alignment Scores:
Pred. No.:   0.025   Length: 110
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 19   Indels: 0
DB: Gaps: 0

US-10-045-180A-4 (1-19) x AAV10895 (1-110)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLaLeuGlnAlaAlaTrpAla 19
          ||||||||||| ::||::||::||| ||||||||||||||||
Db   5 ATGAGGACCCTGCGCCATCCTGCCATTCCTCGTGCTGGCCCTGCAGGCCCAGGCT 61

RESULT 11
ABA75757
ID   ABA75757 standard; DNA; 187 BP.
XX
AC   ABA75757;
XX
DT   01-FEB-2002 (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #24062.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483447/52.
XX

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 4; SEQ ID NO 24062; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 187 BP; 37 A; 60 C; 54 G; 36 T; 0 other;

Alignment Scores:
Pred. No.:   0.0479   Length: 187
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 67.78%   Indels: 0
DB: 22   Gaps: 0

US-10-045-180A-4 (1-19) x ABA75757 (1-187)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          |||||||||||| ::||::::||| |||||||||||||||| |||
Db   13 ATGAGGACCCTGCCATCTCCTTGTGCCATTCTCCTGGTGGGCCCTGCAGGCCCAGGCT 69

RESULT 12
AAK50398
ID AAK50398 standard; DNA; 187 BP.
XX
AC AAK50398;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 24955.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207455.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 24955; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human

CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 187 BP; 37 A; 60 C; 54 G; 36 T; 0 other;

Alignment Scores:
Pred. No.:   0.0479   Length: 187
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 67.78%   Indels: 0
DB: 22   Gaps: 0

US-10-045-180A-4 (1-19) x AAK50398 (1-187)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          |||||||||||| ::||::::||| |||||||||||||||| |||
Db   13 ATGAGGACCCTGCCATCTCCTTGTGCCATTCTCCTGGTGGGCCCTGCAGGCCCAGGCT 69

RESULT 13
ABS23860
ID ABS23860 standard; DNA; 187 BP.
XX
AC ABS23860;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe ORF from lung SEQ ID No 23851.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesis; pulmonary hypertension;
KW hyaline membrane disease; open reading frame; ORF.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 4; SEQ ID No 23851; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614

Left column:

```
CC   probes. Also included are a microarray comprising the novel set of
CC   probes; the novel set of probes which hybridise at high stringency to a
CC   nucleic acid expressed in the human lung; measuring gene expression in a
CC   sample derived from human lung, comprising (a) contacting the array with
CC   a collection of detectably labeled nucleic acids derived from human lung
CC   mRNA, and (b) measuring the label detectably bound to each probe of
CC   the array; identifying exons in a eukaryotic genome, comprising
CC   (a) algorithmically predicting at least one exon from genomic sequences
CC   of the eukaryote; and (b) detecting specific hybridisation of detectably
CC   labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC   having a fragment identical to the predicted exon, the probe is included
CC   in the above mentioned microarray; assigning exons to a single gene,
CC   comprising (a) identifying exons from genomic sequence by the method
CC   above and (b) measuring the expression of each of the exons in several
CC   tissues and/or cell types using hybridisation to a single exon
CC   microarrays having a probe with the exon, where a common pattern of
CC   expression of the exons in the tissues and/or cell types indicates that
CC   the exons should be assigned to a single gene; a peptide comprising one
CC   of 12011 sequences, mentioned in the specification, or encoded by the
CC   probes/open reading frames (ORF). The probes are used for gene
CC   expression analysis, and for identifying exons in a gene, particularly
CC   using human lung derived mRNA and for the study of lung diseases
CC   such as asthma, lung cancer, chronic obstructive pulmonary disease
CC   (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC   fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC   Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC   haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC   pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC   pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC   and hyaline membrane disease. The present sequence is a single exon
CC   probe open reading frame of the invention.
CC   Note: The sequence data for this patent did not form part.
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 187 BP; 37 A; 60 C; 54 G; 36 T; 0 other;

Alignment Scores:
Pred. No.:                0.0479      Length:           187
Score:                   61.00       Matches:           14
Percent Similarity:      84.21%      Conservative:       2
Best Local Similarity:   73.68%      Mismatches:         3
Query Match:             67.78%      Indels:             0
DB:                       24         Gaps:               0

US-10-045-180A-4 (1-19) x ABS23860 (1-187)

Qy        1  MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla  19
             ||||||||||||||||||||||  ::::|||  ||||||||||||||||||||||
Db       13  ATGAGGACCCTCGCCATCCTTGCTGCCATTCCTGTGGCCCTGCAGGCCCAGGCT  69

RESULT 14
AAT26106
ID   AAT26106 standard; cDNA to mRNA; 299 BP.
XX
AC   AAT26106;
XX
DT   18-OCT-1996  (first entry)
XX
DE   Human gene signature HUMGS08342.
XX
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
XX
OS   Homo sapiens.
XX
PN   WO9514772-A1.
XX
PD   01-JUN-1995.
XX
PF   11-NOV-1994;  94WO-JP01916.
```

Right column:

```
XX
PR   12-NOV-1993;  93JP-0355504.
XX
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
XX
PI   Matsubara K, Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 2003; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in AAT19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
XX
SQ   Sequence 299 BP; 67 A; 87 C; 83 G; 61 T; 1 other;

Alignment Scores:
Pred. No.:                0.0853      Length:           299
Score:                   61.00       Matches:           14
Percent Similarity:      84.21%      Conservative:       2
Best Local Similarity:   73.68%      Mismatches:         3
Query Match:             67.78%      Indels:             0
DB:                       16         Gaps:               0

US-10-045-180A-4 (1-19) x AAT26106 (1-299)

Qy        1  MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla  19
             ||||||||||||||||||||||  ::::|||  ||||||||||||||||||||||
Db       18  ATGAGGACCCTCGCCATCCTTGCTGCCATTCCTGTGGCCCTGCAGGCCCAGGCT  74

RESULT 15
AAF20911
ID   AAF20911 standard; DNA; 464 BP.
XX
AC   AAF20911;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human defensin 3 polynucleotide fragment #2478.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
```

XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999; 99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 218; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 464 BP; 110 A; 128 C; 119 G; 107 T; 0 other;

Alignment Scores:
Pred. No.:   0.146   Length: 464
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 67.78%   Indels: 0
DB: 21   Gaps: 0

US-10-045-180A-4 (1-19) x AAF20911 (1-464)

Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        |||||||||| ::||||::|| |||||||||||||||||||||||||||
Db   90 ATGAGGACCCTGCCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCT 146

Search completed: December 17, 2003, 15:06:02
Job time : 72.25 secs

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:       December 17, 2003, 15:06:18 ; Search time 65.6915 Seconds
                                           (without alignments)
                                           963.776 Million cell updates/sec

Title:        US-10-045-180A-4
Perfect score: 90
Sequence:     1 MRTLTLLSAFLLVALQAWA 19

Scoring table: BLOSUM62
              Xgapop  10.0 , Xgapext   0.5
              Ygapop  10.0 , Ygapext   0.5
              Fgapop   6.0 , Fgapext   7.0
              Delop    6.0 , Delext    7.0

Searched:     2211978 seqs, 1666101734 residues

Total number of hits satisfying chosen parameters:        4423956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145026_12635/app_query.fasta_1.860
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10045180 @CGN_1_1_174 @runat_17122003_145026_12635
-NCPU=6 -ICPU=3 -NO_NMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :    Published_Applications NA:*
         1:   /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
         2:   /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
         3:   /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
         4:   /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
         5:   /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
         6:   /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
         7:   /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
         8:   /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
         9:   /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
        10:   /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
        11:   /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
        12:   /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
        13:   /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
        14:   /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
        15:   /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
        16:   /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
        17:   /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
        18:   /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Score	Query Match	Length	DB ID	Description
```
 1   100.0   90   453   14  US-10-013-770-2        Sequence 2, Appli
 2   100.0   90   453   14  US-10-045-180A-2       Sequence 2, Appli
 3   100.0   90  4415   14  US-10-013-770-1        Sequence 1, Appli
 4   100.0   90  4415   14  US-10-045-180A-1       Sequence 1, Appli
 5    67.8   61    57   12  US-10-367-169-45       Sequence 45, Appl
 6    67.8   61    57   13  US-10-076-802-45       Sequence 45, Appl
 7    67.8   61   101   12  US-10-367-169-34       Sequence 34, Appl
 8    67.8   61   101   13  US-10-076-802-34       Sequence 34, Appl
 9    67.8   61   187    9  US-09-864-761-31864    Sequence 31864, A
10    67.8   61   555   13  US-10-252-157-223      Sequence 223, App
11    67.8   61   597    9  US-09-864-761-15342    Sequence 15342, A
12    67.8   61   726   10  US-09-816-828-16       Sequence 16, Appl
13    67.8   61  1348   15  US-10-125-237-78       Sequence 78, Appl
14    67.8   61  1348   15  US-10-105-891-78       Sequence 78, Appl
15    66.7   60   598   13  US-10-027-632-48280    Sequence 48280, A
16    66.7   60   598   13  US-10-027-632-48281    Sequence 48281, A
17    66.7   60   598   13  US-10-027-632-48282    Sequence 48282, A
18    66.7   60   598   13  US-10-027-632-48283    Sequence 48283, A
19    66.7   60   598   13  US-10-027-632-48284    Sequence 48284, A
20    66.7   60   598   14  US-10-027-632-48280    Sequence 48280, A
21    66.7   60   598   14  US-10-027-632-48281    Sequence 48281, A
22    66.7   60   598   14  US-10-027-632-48282    Sequence 48282, A
23    66.7   60   598   14  US-10-027-632-48283    Sequence 48283, A
24    66.7   60   598   14  US-10-027-632-48284    Sequence 48284, A
25    62.2   56   372   11  US-09-918-995-18938    Sequence 18938, A
26    60.0   54   436   15  US-10-171-581-252      Sequence 252, App
27    55.6   50   542   14  US-10-045-180A-8       Sequence 8, Appli
28    55.6   50   654   13  US-10-027-632-141529   Sequence 141529,
29    55.6   50   654   14  US-10-027-632-141529   Sequence 141529,
30    55.6   50   741   13  US-10-027-632-141528   Sequence 141528,
31    55.6   50   741   13  US-10-027-632-141530   Sequence 141530,
32    55.6   50   741   13  US-10-027-632-141528   Sequence 141528,
33    55.6   50   741   14  US-10-027-632-141530   Sequence 141530,
34    55.6   50   860   14  US-10-044-090-473      Sequence 473, App
35    55.6   50  4295   15  US-10-045-180A-7       Sequence 7, Appli
36    54.4   49   366   15  US-10-095-672A-4       Sequence 4, Appli
37    54.4   49   806   14  US-10-044-090-136      Sequence 136, App
38    54.4   49  1375   10  US-09-263-959-289      Sequence 289, App
39    54.4   49 684973   10  US-09-263-959-1        Sequence 1, Appli
40    53.3   48  1182   13  US-10-180-454-1        Sequence 1, Appli
41    53.3   48  1191   11  US-09-892-877-11       Sequence 11, Appl
42    53.3   48  1191   11  US-09-948-783-11       Sequence 11, Appl
43    53.3   48 127197  11  US-09-754-853A-1       Sequence 1, Appli
44    52.2   47   637   13  US-10-027-632-247501   Sequence 247501,
45    52.2   47   637   14  US-10-027-632-247501   Sequence 247501,
```

ALIGNMENTS

```
RESULT 1
US-10-013-770-2
; Sequence 2, Application US/10013770
; Publication No. US20020115151A1
; GENERAL INFORMATION:
;  APPLICANT: GENSET SA
;  TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;                     COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;                     THERAPEUTIC APPLICATIONS
;  NUMBER OF SEQUENCES: 6
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/013,770
```

```
                   FILING DATE: 10-Dec-2001
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 09/486,580
;              FILING DATE: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Hart, Daniel
;              REGISTRATION NUMBER: 40,637
;              REFERENCE/DOCKET NUMBER: GENSET.064C1
;    INFORMATION FOR SEQ ID NO: 2:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 453 BASE PAIRS
;              TYPE: NUCLEOTIDE
;              STRANDEDNESS: DOUBLE
;              TOPOLOGY: LINEAR
;         MOLECULE TYPE: cDNA
;         ORIGINAL SOURCE:
;              ORGANISM: Homo sapiens
;         SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-013-770-2

Alignment Scores:
Pred. No.:             1.71e-07        Length:        453
Score:                 90.00           Matches:       19
Percent Similarity:    100.00%         Conservative:  0
Best Local Similarity: 100.00%         Mismatches:    0
Query Match:           100.00%         Indels:        0
DB:                    14              Gaps:          0

US-10-045-180A-4  (1-19) x US-10-013-770-2  (1-453)

Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52 ATGAGGACCCTCACCCTCTCTGCCTTCCTGCTGGTGGCCCTCAGGCCTGGGCA 108

RESULT 2
US-10-045-180A-2
; Sequence 2, Application US/10045180A
; Publication No. US20020182703A1
; GENERAL INFORMATION:
;    APPLICANT: Bougueleret, Lydie
;    APPLICANT: Chumakov, Ilya
;    TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Compositi
;    TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeutic
;    FILE REFERENCE: GEN-100D1
;    CURRENT APPLICATION NUMBER: US/10/045,180A
;    CURRENT FILING DATE: 2001-10-18
;    PRIOR APPLICATION NUMBER: US 09/486,580
;    PRIOR FILING DATE: 2000-02-25
;    PRIOR APPLICATION NUMBER: PCT/FR98/01864
;    PRIOR FILING DATE: 1998-08-28
;    PRIOR APPLICATION NUMBER: FR 97/10823
;    PRIOR FILING DATE: 1997-08-29
;    NUMBER OF SEQ ID NOS: 12
;    SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;    LENGTH: 453
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (52)..(336)
;    OTHER INFORMATION: Def-X coding sequence
US-10-045-180A-2

Alignment Scores:
Pred. No.:             1.71e-07        Length:        453
Score:                 90.00           Matches:       19
Percent Similarity:    100.00%         Conservative:  0
Best Local Similarity: 100.00%         Mismatches:    0
Query Match:           100.00%         Indels:        0
DB:                    14              Gaps:          0

US-10-045-180A-4  (1-19) x US-10-045-180A-2  (1-453)
```

```
Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52 ATGAGGACCCTCACCCTCTCTGCCTTCCTGCTGGTGGCCCTCAGGCCTGGGCA 108

RESULT 3
US-10-013-770-1
; Sequence 1, Application US/10013770
; Publication No. US20020115151A1
; GENERAL INFORMATION:
;    APPLICANT: GENSET SA
;    TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAc
;                        COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;                        THERAPEUTIC APPLICATIONS
;    NUMBER OF SEQUENCES: 6
;    CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Knobbe, Martens, Olson & Bear
;     STREET: 550 West C Street
;     CITY: San Diego
;     STATE: California
;     COUNTRY: USA
;     ZIP: 92101
;    COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0,  Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/10/013,770
;     FILING DATE: 10-Dec-2001
;    PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 09/486,580
;     FILING DATE: <Unknown>
;    ATTORNEY/AGENT INFORMATION:
;     NAME: Hart, Daniel
;     REGISTRATION NUMBER: 40,637
;     REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 1:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 4415 BASE PAIRS
;     TYPE: NUCLEOTIDE
;     STRANDEDNESS: DOUBLE
;     TOPOLOGY: LINEAR
;    MOLECULE TYPE: DNA
;    ORIGINAL SOURCE:
;     ORGANISM: Homo sapiens
;    FEATURE:
;     NAME/KEY:  Exon 1
;     LOCATION:  1836..1874
;    FEATURE:
;     NAME/KEY:  Exon 2
;     LOCATION:  3394..3577
;    FEATURE:
;     NAME/KEY:  Exon 3
;     LOCATION:  4161..4380
;    FEATURE:
;     NAME/KEY:  start CDS
;     LOCATION:  3406..3408
;    FEATURE:
;     NAME/KEY:  stop CDS
;     LOCATION:  4276..4278
;    FEATURE:
;     NAME/KEY:  polyadenylation site
;     LOCATION:  4374..4379
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-013-770-1

Alignment Scores:
Pred. No.:             3.05e-06        Length:        4415
Score:                 90.00           Matches:       19
Percent Similarity:    100.00%         Conservative:  0
Best Local Similarity: 100.00%         Mismatches:    0
Query Match:           100.00%         Indels:        0
```

```
DB:                    14      Gaps:        0

US-10-045-180A-4 (1-19) x US-10-013-770-1 (1-4415)

Qy       1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3406 ATGAGGACCCTCACCCTCCTCTCTGCCTTCTCCTGGTGGCCCTTCAGGCCTGGGCA 3462

RESULT 4
US-10-045-180A-1
; Sequence 1, Application US/10045180A
; Publication No. US20020182703A1
GENERAL INFORMATION:
;  APPLICANT: Bougueleret, Lydie
;  APPLICANT: Chumakov, Ilya
;  TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Compositi
;  TITLE OF INVENTION:  Containing Them and Applications to Diagnosis and to Therapeutic
;  FILE REFERENCE: GEN-100D1
;  CURRENT APPLICATION NUMBER: US/10/045,180A
;  CURRENT FILING DATE:  2001-10-18
;  PRIOR APPLICATION NUMBER: US 09/486,580
;  PRIOR FILING DATE: 2000-02-25
;  PRIOR APPLICATION NUMBER: PCT/FR98/01864
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: FR 97/10823
;  PRIOR FILING DATE: 1997-08-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.1
;
;  SEQ ID NO 1
;    LENGTH: 4415
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (1)..(4415)
;      OTHER INFORMATION: Def-X genomic sequence
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (85)..(85)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (143)..(143)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (670)..(670)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (970)..(970)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (1111)..(1111)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (1150)..(1150)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: CAAT_signal
;      LOCATION: (1711)..(1714)
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY: TATA_signal
;      LOCATION: (1758)..(1767)
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (1780)..(1780)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (1836)..(1874)
;      OTHER INFORMATION: Exon 1
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (1875)..(1880)
;      OTHER INFORMATION: splice donor site
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (1974)..(1974)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (2117)..(2117)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (2133)..(2133)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (2155)..(2335)
;      OTHER INFORMATION: Alu insertion
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (2186)..(2186)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (2191)..(2191)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (2367)..(2367)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (2710)..(2780)
;      OTHER INFORMATION: L1 fragment insertion
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (3391)..(3393)
;      OTHER INFORMATION: splice acceptor site
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (3394)..(3577)
;      OTHER INFORMATION: Exon 2
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (3406)..(3408)
;      OTHER INFORMATION: Translation initiation codon (ATG)
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (3578)..(3583)
;      OTHER INFORMATION: splice donor site
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (4123)..(4123)
;      OTHER INFORMATION: n = a, c, g, or t.
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (4161)..(4163)
;      OTHER INFORMATION: splice acceptor site
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (4164)..(4379)
;      OTHER INFORMATION: Exon 3
;    FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (4274)..(4276)
;      OTHER INFORMATION: Translation termination codon (TAA)
;    FEATURE:
;      NAME/KEY: polyA_signal
```

```
;   LOCATION: (4374)..(4379)
;   OTHER INFORMATION:
US-10-045-180A-1

Alignment Scores:
Pred. No.:                  3.05e-06        Length:            4415
Score:                      90.00          Matches:           19
Percent Similarity:         100.00%        Conservative:      0
Best Local Similarity:      100.00%        Mismatches:        0
Query Match:                100.00%        Indels:            0
DB:                         14             Gaps:              0

US-10-045-180A-4 (1-19) x US-10-045-180A-1 (1-4415)

Qy     1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3406 ATGAGGAGGACCCTCACCCTCCTCTGCCTTTCTCCTGTGGTGGCCCTTCAGGCCTGGGCA 3462

RESULT 5
US-10-367-169-45
; Sequence 45, Application US/10367169
; Publication No. US20030228660A1
; GENERAL INFORMATION:
; APPLICANT: Gray, Jeff
; APPLICANT: Buechler, Joe
; APPLICANT: Veeramallu, Uday Kumar
; TITLE OF INVENTION: EUKARYOTIC SIGNAL SEQUENCES FOR POLYPEPTIDE EXPRESSION AND POLYPE
; FILE REFERENCE: 11055US02
; CURRENT APPLICATION NUMBER: US/10/367,169
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: 10/076,802
; PRIOR FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 57
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Neutrophil defensin 1,2,3 signal sequence
US-10-367-169-45

Alignment Scores:
Pred. No.:                  0.00319        Length:            57
Score:                      61.00          Matches:           14
Percent Similarity:         84.21%         Conservative:      2
Best Local Similarity:      73.68%         Mismatches:        3
Query Match:                67.78%         Indels:            0
DB:                         12             Gaps:              0

US-10-045-180A-4 (1-19) x US-10-367-169-45 (1-57)

Qy     1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
           ||||||||||  :::||:::||       |||
Db     1 ATGCGGTACTCTGGCTATCCTTGCAGCTATTCTGCTTGTTGCACTGCAGGCTCAAGCG 57

RESULT 6
US-10-076-802-45
; Sequence 45, Application US/10076802
; Publication No. US20030162249A1
; GENERAL INFORMATION:
; APPLICANT: Gray, Jeff
; APPLICANT: Buechler, Joe
; APPLICANT: Veeramallu, Uday K
; TITLE OF INVENTION: EUKARYOTIC SIGNAL SEQUENCES FOR PROKARYOTIC EXPRESSION
; FILE REFERENCE: 014907-004700US
; CURRENT APPLICATION NUMBER: US/10/076,802
; CURRENT FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 57
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Neutrophil defensin 1,2,3 signal sequence
US-10-076-802-45

Alignment Scores:
Pred. No.:                  0.00319        Length:            57
Score:                      61.00          Matches:           14
Percent Similarity:         84.21%         Conservative:      2
Best Local Similarity:      73.68%         Mismatches:        3
Query Match:                67.78%         Indels:            0
DB:                         13             Gaps:              0

US-10-045-180A-4 (1-19) x US-10-076-802-45 (1-57)

Qy     1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
           ||||||||||  :::||:::||       |||
Db     1 ATGCGGTACTCTGGCTATCCTTGCAGCTATTCTGCTTGTTGCACTGCAGGCTCAAGCG 57

RESULT 7
US-10-367-169-34/c
; Sequence 34, Application US/10367169
; Publication No. US20030228660A1
; GENERAL INFORMATION:
; APPLICANT: Gray, Jeff
; APPLICANT: Buechler, Joe
; APPLICANT: Veeramallu, Uday Kumar
; TITLE OF INVENTION: EUKARYOTIC SIGNAL SEQUENCES FOR POLYPEPTIDE EXPRESSION AND POLYPI
; TITLE OF INVENTION: DISPLAY LIBRARIES
; FILE REFERENCE: 11055US02
; CURRENT APPLICATION NUMBER: US/10/367,169
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: 10/076,802
; PRIOR FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 101
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer EE
US-10-367-169-34

Alignment Scores:
Pred. No.:                  0.00658        Length:            101
Score:                      61.00          Matches:           14
Percent Similarity:         84.21%         Conservative:      2
Best Local Similarity:      73.68%         Mismatches:        3
Query Match:                67.78%         Indels:            0
DB:                         12             Gaps:              0

US-10-045-180A-4 (1-19) x US-10-367-169-34 (1-101)

Qy     1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
           ||||||||||  :::||:::||       |||
Db    77 ATGCGGTACTCTGGCTATCCTTGCAGCTATTCTGCTTGTTGCACTGCAGGCTCAAGCG 101

RESULT 8
US-10-076-802-34/c
; Sequence 34, Application US/10076802
; Publication No. US20030162249A1
; GENERAL INFORMATION:
; APPLICANT: Gray, Jeff
; APPLICANT: Buechler, Joe
; APPLICANT: Veeramallu, Uday K
; TITLE OF INVENTION: EUKARYOTIC SIGNAL SEQUENCES FOR PROKARYOTIC EXPRESSION
; FILE REFERENCE: 014907-004700US
; CURRENT APPLICATION NUMBER: US/10/076,802
; CURRENT FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 34
; LENGTH: 101
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer EE
US-10-076-802-34

Alignment Scores:
Pred. No.:              0.00658          Length:         101
Score:                  61.00            Matches:        14
Percent Similarity:     84.21%           Conservative:   2
Best Local Similarity:  73.68%           Mismatches:     3
Query Match:            67.78%           Indels:         0
DB:                                      Gaps:           0

US-10-045-180A-4 (1-19) x US-10-076-802-34 (1-101)

Qy    1 MetArgThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          |||||||   :::||:::||| |||||||||||||
Db   77 ATGCGTACTCTGGCTATCCTTGCAGCTATTCTGCTTGTTGCACTGCAGGCTCAAGCG 21

RESULT 9
US-09-864-761-31864
; Sequence 31864, Application US/09864761
; Patent No. US2002004876341
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
```

```
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31864
; LENGTH: 187
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF233439.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6e+02
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 18
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 24
; OTHER INFORMATION: SWISSPROT HIT: P11479, EVALUE 7.00e-17
; OTHER INFORMATION: NT HIT: X52053.1, EVALUE 1.00e-101
; OTHER INFORMATION: EST_HUMAN HIT: AA321316.1, EVALUE 1.00e-101
US-09-864-761-31864

Alignment Scores:
Pred. No.:              0.0143           Length:         187
Score:                  61.00            Matches:        14
Percent Similarity:     84.21%           Conservative:   2
Best Local Similarity:  73.68%           Mismatches:     3
Query Match:            67.78%           Indels:         0
DB:                     9                Gaps:           0

US-10-045-180A-4 (1-19) x US-09-864-761-31864 (1-187)

Qy    1 MetArgThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          ||||||||   :::||:::|| |||||||||||||||||||||
Db   13 ATGAGGACCCTGCGCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCT 69

RESULT 10
US-10-252-157-223
; Sequence 223, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Pearson, Cecelia I.
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252,157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 223
; LENGTH: 555
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030190640A1 903338.12
US-10-252-157-223

Alignment Scores:
Pred. No.:              0.0568           Length:         555
Score:                  61.00            Matches:        14
Percent Similarity:     84.21%           Conservative:   2
Best Local Similarity:  73.68%           Mismatches:     3
Query Match:            67.78%           Indels:         0
DB:                     13               Gaps:           0

US-10-045-180A-4 (1-19) x US-10-252-157-223 (1-555)

Qy    1 MetArgThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          ||||||||   :::||:::|| |||||||||||||||||||||
Db  145 ATGAGGACCCTGCGCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCT 201

RESULT 11
US-09-864-761-15342
; Sequence 15342, Application US/09864761
; Patent No. US2002004876341
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
```

; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15342
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF233439.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6e+02
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 18
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 24
US-09-864-761-15342

Alignment Scores:
Pred. No.:   0.0623   Length: 597
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 67.78%   Indels: 0
DB: 9   Gaps: 0

US-10-045-180A-4 (1-19) x US-09-864-761-15342 (1-597)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
         |||||||||| :::|||| ::||||||||||||||||||||
Db 364 ATGAGGACCCTGCCATCCTTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCT 420

RESULT 12
US-09-816-828-16

; Sequence 16, Application US/0916828
; Patent No. US20020150898A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhao, Qing A.
; APPLICANT: Zhang, Jie
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020150898A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 791CIP2B
; CURRENT APPLICATION NUMBER: US/09/816,828
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/770,160
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/552,929
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: pt_Fl_genes Version 2.0
; SEQ ID NO 16
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (151)..(522)
US-09-816-828-16

Alignment Scores:
Pred. No.:   0.0797   Length: 726
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 67.78%   Indels: 0
DB: 10   Gaps: 0

US-10-045-180A-4 (1-19) x US-09-816-828-16 (1-726)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
         |||||||||| :::|||| ::||||||||||||||||||||
Db 241 ATGAGGACCCTGCCATCCTTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCT 297

RESULT 13
US-10-125-237-78

; Sequence 78, Application US/10125237
; Publication No. US20030022329A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhao, Qing A.
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhang, Jie
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030022329A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 791CIP2ADIV
; CURRENT APPLICATION NUMBER: US/10/125,237
; CURRENT FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 09/668,317
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/552,929

```
;   PRIOR FILING DATE: 2000-04-18
;   NUMBER OF SEQ ID NOS: 91
;   SOFTWARE: pt_FL_genes Version 2.0
;   SEQ ID NO 78
;   LENGTH: 1348
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1143)
US-10-125-237-78

Alignment Scores:
Pred. No.:              0.174          Length:         1348
Score:                  61.00          Matches:          14
Percent Similarity:     84.21%         Conservative:      2
Best Local Similarity:  73.68%         Mismatches:        3
Query Match:            67.78%         Indels:            0
DB:                     15             Gaps:              0

US-10-045-180A-4 (1-19) x US-10-125-237-78 (1-1348)

Qy    1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        |||||||||||   ::|||:::|||   |||||||||||||||||||||
Db  859 ATGAGGACCCTGCCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCT 915


RESULT 14
US-10-105-891-78
; Sequence 78, Application US/10105891
; Publication No. US20030073099A1
; GENERAL INFORMATION:
;   APPLICANT: Tang, Y. Tom
;   APPLICANT: Liu, Chenghua
;   APPLICANT: Zhou, Ping
;   APPLICANT: Asundi, Vinod
;   APPLICANT: Ren, Feiyan
;   APPLICANT: Zhao, Qing A.
;   APPLICANT: Xue, Aidong J.
;   APPLICANT: Zhang, Jie
;   APPLICANT: Wehrman, Tom
;   APPLICANT: Wang, Jian-Rui
;   APPLICANT: Drmanac, Radoje T.
;   TITLE OF INVENTION: No. US20030073099Alel Nucleic Acids and
;   TITLE OF INVENTION: Polypeptides
;   FILE REFERENCE: 791CIP2A
;   CURRENT APPLICATION NUMBER: US/10/105,891
;   CURRENT FILING DATE: 2002-03-25
;   PRIOR APPLICATION NUMBER: 09/668,317
;   PRIOR FILING DATE: 2000-09-22
;   PRIOR APPLICATION NUMBER: 09/552,929
;   PRIOR FILING DATE: 2000-04-18
;   NUMBER OF SEQ ID NOS: 91
;   SOFTWARE: pt_FL_genes Version 2.0
;   SEQ ID NO 78
;   LENGTH: 1348
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1143)
US-10-105-891-78

Alignment Scores:
Pred. No.:              0.174          Length:         1348
Score:                  61.00          Matches:          14
Percent Similarity:     84.21%         Conservative:      2
Best Local Similarity:  73.68%         Mismatches:        3
Query Match:            67.78%         Indels:            0
DB:                     15             Gaps:              0

US-10-045-180A-4 (1-19) x US-10-105-891-78 (1-1348)

Qy    1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
```

```
Db  859 ATGAGGACCCTGCCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCAGGCT 915


RESULT 15
US-10-027-632-48280
; Sequence 48280, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   TITLE OF INVENTION: Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 48280
;   LENGTH: 598
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-48280

Alignment Scores:
Pred. No.:              0.0959         Length:          598
Score:                  60.00          Matches:          14
Percent Similarity:     84.21%         Conservative:      2
Best Local Similarity:  73.68%         Mismatches:        3
Query Match:            66.67%         Indels:            0
DB:                     13             Gaps:              0

US-10-045-180A-4 (1-19) x US-10-027-632-48280 (1-598)

Qy    1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        |||||||||||   ::|||:::|||
Db  402 ATGAGGACCCTGCCATCCTTGCTGCCMTTCTCCTGGTGGCCMTTCTCCTGGTGGCCCTGCAGGCCCAGGCT 458


Search completed: December 17, 2003, 21:35:59
Job time : 67.6915 secs
```

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      December 17, 2003, 14:50:33 ; Search time 19 Seconds
                                  (without alignments)
                                  441.383 Million cell updates/sec

Title:       US-10-045-180A-4
Perfect score: 90
Sequence:    1 MRTLTLLSAFLIVALQAWA 19

Scoring table: BLOSUM62
             Xgapop  10.0 , Xgapext   0.5
             Ygapop  10.0 , Ygapext   0.5
             Fgapop   6.0 , Fgapext   7.0
             Delop    6.0 , Delext    7.0

Searched:    569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters:     1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool_p/US10045180/runat_17122003_145024_12553/app_query.fasta_1.860
-DB=Issued_PatentS_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=Bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10045180 @CGN_1_193 @runat_17122003_145024_12553 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Issued_Patents NA:*
         1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
         2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
         3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
         4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
         5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
         6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

                     %
Result         Query
No.    Score   Match  Length DB  ID                      Description
-----  -----   -----  ------ --  --                      -----------
  1      90    100.0    453   4  US-09-486-580A-2         Sequence 2, Appli
  2      90    100.0   4415   4  US-09-486-580A-1         Sequence 1, Appli
  3      64     71.1    452   1  US-08-158-189-6          Sequence 6, Appli
  4      64     71.1    452   4  US-09-016-434-1198       Sequence 1198, Ap
  5      63     70.0    195   1  US-08-158-189-9          Sequence 9, Appli
  6      63     70.0    202   1  US-08-158-189-8          Sequence 8, Appli
  7      61     67.8    110   1  US-08-575-052-9          Sequence 9, Appli
  8      61     67.8    110   1  US-08-614-516A-9         Sequence 9, Appli
  9      61     67.8    110   1  US-08-770-557-9          Sequence 9, Appli
 10      61     67.8    110   1  US-08-405-234-9          Sequence 9, Appli
 11      61     67.8    203   1  US-08-158-189-10         Sequence 10, Appli
 12      61     67.8    203   1  US-08-158-189-11         Sequence 11, Appli
```

```
 13      61     67.8    218   1  US-08-158-189-38
 14      59     65.6    424   1  US-08-158-189-4          Sequence 38, Appl
 15      59     65.6   1097   1  US-08-021-608D-5         Sequence 4, Appli
 16    c 59     65.6   1097   1  US-08-726-160-5          Sequence 5, Appli
 17    c 59     65.6   1097   5  PCT-US94-01782-5         Sequence 5, Appli
 18    c 59     65.6   2880   1  US-08-158-189-1          Sequence 1, Appli
 19      58     64.4     92   1  US-08-575-052-12         Sequence 12, Appli
 20    c 58     64.4     92   1  US-08-614-516A-12        Sequence 12, Appli
 21    c 58     64.4     92   1  US-08-770-557-12         Sequence 12, Appli
 22    c 58     64.4     92   1  US-08-405-234-12         Sequence 12, Appli
 23      56     62.2    105   3  US-09-042-071-44         Sequence 44, Appl
 24    c 52     57.8   5687   2  US-08-380-403A-3         Sequence 3, Appli
 25    c 52     57.8   5687   2  US-08-895-628-3          Sequence 3, Appli
 26    c 52     57.8   5687   2  US-08-895-810D-3         Sequence 3, Appli
 27      51     56.7    445   1  US-08-486-013-66         Sequence 66, Appli
 28      51     56.7    445   2  US-08-482-279-66         Sequence 66, Appli
 29      51     56.7    445   2  US-08-342-268-66         Sequence 66, Appl
 30      51     56.7    445   3  US-09-015-968-66         Sequence 66, Appl
 31      51     56.7    445   4  US-09-397-386-66         Sequence 66, Appl
 32      51     56.7   2408   1  US-08-486-013-69         Sequence 69, Appl
 33      51     56.7   2408   2  US-08-482-279-69         Sequence 69, Appl
 34      51     56.7   2408   2  US-08-342-268-69         Sequence 69, Appl
 35      51     56.7   2408   3  US-09-015-968-69         Sequence 69, Appl
 36      51     56.7   2408   4  US-09-397-386-69         Sequence 69, Appl
 37      49     54.4    366   1  US-08-423-383-5          Sequence 5, Appli
 38      49     54.4    366   2  US-08-437-353A-5         Sequence 5, Appl
 39      49     54.4    377   1  US-08-486-013-48         Sequence 48, Appl
 40      49     54.4    377   2  US-08-482-279-48         Sequence 48, Appl
 41      49     54.4    377   2  US-08-342-268-48         Sequence 48, Appl
 42      49     54.4    377   3  US-09-015-968-48         Sequence 48, Appl
 43      49     54.4    377   4  US-09-397-386-48         Sequence 48, Appl
 44      49     54.4    388   1  US-08-486-013-52         Sequence 52, Appl
 45      49     54.4    388   2  US-08-482-279-52         Sequence 52, Appl
```

```
                              ALIGNMENTS

RESULT 1
US-09-486-580A-2
; Sequence 2, Application US/09486580A
; Patent No. 6329340
; GENERAL INFORMATION:
;  APPLICANT: GENSET SA
;  TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;  TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;  TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
;  NUMBER OF SEQUENCES: 6
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/486,580A
;   FILING DATE: FEBRUARY 25, 2000
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
;  INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH: 453 BASE PAIRS
;    TYPE: NUCLEOTIDE
;    STRANDEDNESS: DOUBLE
;    TOPOLOGY: LINEAR
;    MOLECULE TYPE: cDNA
```

```
;   ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
US-09-486-580A-2

Alignment Scores:
Pred. No.:              3.49e-07      Length:          453
Score:                  90.00        Matches:          19
Percent Similarity:     100.00%      Conservative:      0
Best Local Similarity:  100.00%      Mismatches:        0
Query Match:            100.00%      Indels:            0
DB:                     4            Gaps:              0

US-10-045-180A-4 (1-19) x US-09-486-580A-2 (1-453)

Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52 ATGAGGACCCTCACCCTCCTCTCGCCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCA 108

RESULT 2
US-09-486-580A-1
; Sequence 1, Application US/09486580A
; Patent No. 6329340
; GENERAL INFORMATION:
;   APPLICANT: GENSET SA
;   TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;   TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;   TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Knobbe, Martens, Olson & Bear
;     STREET: 550 West C Street
;     CITY: San Diego
;     STATE: California
;     COUNTRY: USA
;     ZIP: 92101
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/486,580A
;     FILING DATE: FEBRUARY 25, 2000
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Hart, Daniel
;     REGISTRATION NUMBER: 40,637
;     REFERENCE/DOCKET NUMBER: GENSET.064C1
;   INFORMATION FOR SEQ ID NO: 1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 4415 BASE PAIRS
;       TYPE: NUCLEOTIDE
;       STRANDEDNESS: DOUBLE
;       TOPOLOGY: LINEAR
;     MOLECULE TYPE: DNA
;     ORIGINAL SOURCE:
;       ORGANISM: Homo sapiens
;     FEATURE:
;       NAME/KEY: Exon 1
;       LOCATION: 1836..1874
;     FEATURE:
;       NAME/KEY: Exon 2
;       LOCATION: 3394..3577
;     FEATURE:
;       NAME/KEY: Exon 3
;       LOCATION: 4161..4380
;     FEATURE:
;       NAME/KEY: start CDS
;       LOCATION: 3406..3408
;     FEATURE:
;       NAME/KEY: stop CDS
;       LOCATION: 4276..4278
;     FEATURE:
;       NAME/KEY: polyadenylation site
;       LOCATION: 4374..4379
US-09-486-580A-1

Alignment Scores:
Pred. No.:              5.33e-06      Length:         4415
Score:                  90.00        Matches:          19
Percent Similarity:     100.00%      Conservative:      0
Best Local Similarity:  100.00%      Mismatches:        0
Query Match:            100.00%      Indels:            0
DB:                     4            Gaps:              0

US-10-045-180A-4 (1-19) x US-09-486-580A-1 (1-4415)

Qy    1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3406 ATGAGGACCCTCACCCTCCTCTCGCCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCA 3462

RESULT 3
US-08-158-189-6
; Sequence 6, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
;   APPLICANT: Bevins, Charles L.
;   APPLICANT: Jones, Douglas E.
;   TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
;   TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
;   NUMBER OF SEQUENCES: 51
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
;     STREET: One Liberty Place - 46th Floor
;     CITY: Philadelphia
;     STATE: PA
;     COUNTRY: U.S.A.
;     ZIP: 19103
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/158,189
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/07/888,232
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Johnson, Philip S.
;     REGISTRATION NUMBER: 27,200
;     REFERENCE/DOCKET NUMBER: CH-0219
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 215-568-3100
;     TELEFAX: 215-568-3439
;   INFORMATION FOR SEQ ID NO: 6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 452 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: unknown
;     FEATURE:
;       NAME/KEY: CDS
;       LOCATION: 19..321
US-08-158-189-6

Alignment Scores:
Pred. No.:              0.0117        Length:          452
Score:                  64.00        Matches:          15
Percent Similarity:     89.47%       Conservative:      2
Best Local Similarity:  78.95%       Mismatches:        2
Query Match:            71.11%       Indels:            0
DB:                     1            Gaps:              0

US-10-045-180A-4 (1-19) x US-08-158-189-6 (1-452)
```

Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          |||||||||||||||||||:||:|||:||||| ||||||||||||||||||||||
Db   19 ATGAGAACCCTCACCATCCTCACTGCTGTTCCTCCTGTGGCCCTCCAGGCCAAGGCT 75

RESULT 4
US-09-016-434-1198
; Sequence 1198, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;   STREET: 3174 PORTER DRIVE
;   CITY: PALO ALTO
;   STATE: CALIFORNIA
;   COUNTRY: USA
;   ZIP: 94304
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/016,434
;   FILING DATE: HEREWITH
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Zeller, Karen J.
;   REGISTRATION NUMBER: 37,071
;   REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (650) 855-0555
;   TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1198:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 452 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: GENBANK
;   CLONE: g1181546
US-09-016-434-1198

Alignment Scores:
Pred. No.:   0.0117   Length: 452
Score: 64.00   Matches: 15
Percent Similarity: 89.47%   Conservative: 2
Best Local Similarity: 78.95%   Mismatches: 2
Query Match: 71.11%   Indels: 0
DB: Gaps: 0

US-10-045-180A-4 (1-19) x US-09-016-434-1198 (1-452)

Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          |||||||||||||||||||:||:|||:||||| ||||||||||||||||||||||
Db   19 ATGAGAACCCTCACCATCCTCACTGCTGTTCCTCCTGTGGCCCTCCAGGCCAAGGCT 75

RESULT 5
US-08-158-189-9
; Sequence 9, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
; APPLICANT: Bevins, Charles L.

APPLICANT: Jones, Douglas E.
TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
 STREET: One Liberty Place - 46th Floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: U.S.A.
 ZIP: 19103
COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/158,189
 FILING DATE:
 CLASSIFICATION: 435
PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/888,232
 FILING DATE:
ATTORNEY/AGENT INFORMATION:
 NAME: Johnson, Philip S.
 REGISTRATION NUMBER: 27,200
 REFERENCE/DOCKET NUMBER: CH-0219
TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-568-3100
 TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
  LENGTH: 195 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: unknown
US-08-158-189-9

Alignment Scores:
Pred. No.:   0.00641   Length: 195
Score: 63.00   Matches: 15
Percent Similarity: 84.21%   Conservative: 1
Best Local Similarity: 78.95%   Mismatches: 3
Query Match: 70.00%   Indels: 0
DB: 1   Gaps: 0

US-10-045-180A-4 (1-19) x US-08-158-189-9 (1-195)

Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
          |||||||||||||||||||:||:|||:|||||||| ||||||||||||||||||||
Db   70 ATGAGGACCCTCGCTCTGCCATTCCTGCTGCCTGTCCTGTGGGGCCCTGCAGGCCCAGGCT 126

RESULT 6
US-08-158-189-8
; Sequence 8, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
; APPLICANT: Bevins, Charles L.
; APPLICANT: Jones, Douglas E.
; TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
;   STREET: One Liberty Place - 46th.Floor
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: U.S.A.
;   ZIP: 19103
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS

```
     SOFTWARE:  PatentIn Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/158,189
     FILING DATE:
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  US/07/888,232
     FILING DATE:
   ATTORNEY/AGENT INFORMATION:
     NAME:  Johnson, Philip S.
     REGISTRATION NUMBER:  27,200
     REFERENCE/DOCKET NUMBER:  CH-0219
   TELECOMMUNICATION INFORMATION:
     TELEPHONE:  215-568-3100
     TELEFAX:  215-568-3439
 INFORMATION FOR SEQ ID NO:  8:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  202 base pairs
     TYPE:  nucleic acid
     STRANDEDNESS:  double
     TOPOLOGY:  unknown
 US-08-158-189-8

 Alignment Scores:
 Pred. No.:                0.00669          Length:         202
 Score:                   63.00            Matches:         15
 Percent Similarity:      84.21%           Conservative:     1
 Best Local Similarity:   78.95%           Mismatches:       3
 Query Match:             70.00%           Indels:           0
 DB:                      1                Gaps:             0

 US-10-045-180A-4 (1-19) x US-08-158-189-8 (1-202)

 Qy    1  MetArgThrLeuThrLeuSerAlaPheLeuValAlaAlaLeuGlnAlaTrpAla 19
          ||||||||||||||||  ::||||||||||||||||||  |||||||||||
 Db   77  ATGAGGACCCTGCCTGCTTGCGCCATTCCTGCTGTGGCCTGAGGCCCAGGCT 133

 RESULT 7
 US-08-575-052-9
 ; Sequence 9, Application US/08575052
 ; Patent No. 5593866
 ; GENERAL INFORMATION:
 ;   APPLICANT:  HANCOCK, ROBERT E.W.
 ;   APPLICANT:  PIERS, KEVIN L.
 ;   APPLICANT:  BROWN, MELISSA H.
 ;   TITLE OF INVENTION:  CATIONIC PEPTIDES AND METHOD FOR PRODUCTION
 ;   NUMBER OF SEQUENCES:  26
 ;   CORRESPONDENCE ADDRESS:
 ;     ADDRESSEE:  SPENSLEY HORN JUBAS & LUBITZ
 ;     STREET:  1880 CENTURY PARK EAST, FIFTH FLOOR
 ;     CITY:  LOS ANGELES
 ;     STATE:  CALIFORNIA
 ;     COUNTRY:  U.S.A.
 ;     ZIP:  90067
 ;   COMPUTER READABLE FORM:
 ;     MEDIUM TYPE:  Floppy disk
 ;     COMPUTER:  IBM PC compatible
 ;     OPERATING SYSTEM:  PC-DOS/MS-DOS
 ;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
 ;   CURRENT APPLICATION DATA:
 ;     APPLICATION NUMBER:  US/08/575,052
 ;     FILING DATE:
 ;     CLASSIFICATION:  435
 ;   PRIOR APPLICATION DATA:
 ;     APPLICATION NUMBER:  US/08/110,502
 ;     FILING DATE:  20-AUG-1993
 ;   ATTORNEY/AGENT INFORMATION:
 ;     NAME:  WETHERELL, JR., PH.D., JOHN R.
 ;     REGISTRATION NUMBER:  31,678
 ;     REFERENCE/DOCKET NUMBER:  PD2823
 ;   TELECOMMUNICATION INFORMATION:
 ;     TELEPHONE:  619/455-5100
 ;     TELEFAX:  619/455-5110
```

```
 INFORMATION FOR SEQ ID NO:  9:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  110 base pairs
     TYPE:  nucleic acid
     STRANDEDNESS:  single
     TOPOLOGY:  linear
   MOLECULE TYPE:  DNA (genomic)
   FEATURE:
     NAME/KEY:  CDS
     LOCATION:  2..110
 US-08-575-052-9

 Alignment Scores:
 Pred. No.:                0.00721          Length:         110
 Score:                   61.00            Matches:         14
 Percent Similarity:      84.21%           Conservative:     2
 Best Local Similarity:   73.68%           Mismatches:       3
 Query Match:             67.78%           Indels:           0
 DB:                      1                Gaps:             0

 US-10-045-180A-4 (1-19) x US-08-575-052-9 (1-110)

 Qy    1  MetArgThrLeuThrLeuSerAlaPheLeuValAlaAlaLeuGlnAlaTrpAla 19
          ||||||||||||||||  :::||||||  ||||||||||||  |||||||||
 Db    5  ATGAGGACCCTGCCATCCTTGCGCCATTCCTCCTGTGGCCTGAGGCCCAGGCT 61

 RESULT 8
 US-08-614-516A-9
 ; Sequence 9, Application US/08614516A
 ; Patent No. 5688767
 ; GENERAL INFORMATION:
 ;   APPLICANT:  HANCOCK, ROBERT E.W.
 ;   APPLICANT:  PIERS, KEVIN L.
 ;   APPLICANT:  BROWN, MELISSA H.
 ;   APPLICANT:  KELLY, NIAMH
 ;   TITLE OF INVENTION:  TREATMENT OF ENDOTOXIN-ASSOCIATED
 ;   TITLE OF INVENTION:  DISORDERS WITH CATIONIC PEPTIDES
 ;   NUMBER OF SEQUENCES:  26
 ;   CORRESPONDENCE ADDRESS:
 ;     ADDRESSEE:  FISH & RICHARDSON P.C.
 ;     STREET:  4225 Executive Square, Ste. 1400
 ;     CITY:  La Jolla
 ;     STATE:  CALIFORNIA
 ;     COUNTRY:  U.S.A.
 ;     ZIP:  92037
 ;   COMPUTER READABLE FORM:
 ;     MEDIUM TYPE:  Floppy disk
 ;     COMPUTER:  IBM PC compatible
 ;     OPERATING SYSTEM:  PC-DOS/MS-DOS
 ;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
 ;   CURRENT APPLICATION DATA:
 ;     APPLICATION NUMBER:  US/08/614,516A
 ;     FILING DATE:  13-MAR-1996
 ;     CLASSIFICATION:  435
 ;   ATTORNEY/AGENT INFORMATION:
 ;     NAME:  HAILE, PH.D., LISA A.
 ;     REGISTRATION NUMBER:  38,347
 ;     REFERENCE/DOCKET NUMBER:  07420/010001
 ;   TELECOMMUNICATION INFORMATION:
 ;     TELEPHONE:  619/678-5070
 ;     TELEFAX:  619/678-5099
 ; INFORMATION FOR SEQ ID NO:  9:
 ;   SEQUENCE CHARACTERISTICS:
 ;     LENGTH:  110 base pairs
 ;     TYPE:  nucleic acid
 ;     STRANDEDNESS:  single
 ;     TOPOLOGY:  linear
 ;   MOLECULE TYPE:  DNA (genomic)
 ;   FEATURE:
 ;     NAME/KEY:  CDS
 ;     LOCATION:  2..110
 US-08-614-516A-9
```

```
Alignment Scores:
 Pred. No.:              0.00721    Length:           110
 Score:                  61.00      Matches:           14
 Percent Similarity:     84.21%     Conservative:       2
 Best Local Similarity:  73.68%     Mismatches:         3
 Query Match:            67.78%     Indels:             0
 DB:                     1          Gaps:               0

US-10-045-180A-4 (1-19) x US-08-614-516A-9 (1-110)

Qy  1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        ||||||||||||     :::||||::|||  ||||||||||||||||||| |||
Db  5 ATGAGGACCCTGCGCCATCCTGCTGCTGGTGGCCCTGCAGGCCCAGGCT 61

RESULT 9
US-08-770-557-9
; Sequence 9, Application US/08770557
; Patent No. 5707855
; GENERAL INFORMATION:
;   APPLICANT: HANCOCK, ROBERT E.W.
;   APPLICANT: PIERS, KEVIN L.
;   APPLICANT: BROWN, MELISSA H.
;   TITLE OF INVENTION: CATIONIC PEPTIDES AND METHOD FOR PRODUCTION
;   NUMBER OF SEQUENCES: 26
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
;     STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
;     CITY: LOS ANGELES
;     STATE: CALIFORNIA
;     COUNTRY: U.S.A.
;     ZIP: 90067
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/770,557
;     FILING DATE: 20-DEC-1996
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/08/575,052
;     FILING DATE:
;     APPLICATION NUMBER: US/08/110,502
;     FILING DATE: 20-AUG-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: WETHERELL, JR., PH.D., JOHN R.
;     REGISTRATION NUMBER: 31,678
;     REFERENCE/DOCKET NUMBER: PD2823
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 619/455-5100
;     TELEFAX: 619/455-5110
; INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 110 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 2..110
US-08-770-557-9

Alignment Scores:
 Pred. No.:              0.00721    Length:           110
 Score:                  61.00      Matches:           14
 Percent Similarity:     84.21%     Conservative:       2
 Best Local Similarity:  73.68%     Mismatches:         3
 Query Match:            67.78%     Indels:             0
 DB:                     1          Gaps:               0

US-10-045-180A-4 (1-19) x US-08-770-557-9 (1-110)

Qy  1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        ||||||||||||     :::||||::|||  ||||||||||||||||||| |||
Db  5 ATGAGGACCCTGCGCCATCCTGCTGCTGGTGGCCCTGCAGGCCCAGGCT 61

RESULT 10
US-08-405-234-9
; Sequence 9, Application US/08405234
; Patent No. 5789377
; GENERAL INFORMATION:
;   APPLICANT: HANCOCK, ROBERT E.W.
;   APPLICANT: PIERS, KEVIN L.
;   APPLICANT: BROWN, MELISSA H.
;   APPLICANT: KELLY, NIAMH
;   TITLE OF INVENTION: TREATMENT OF ENDOTOXIN-ASSOCIATED DISORDERS
;   TITLE OF INVENTION: WITH CATIONIC PEPTIDES
;   NUMBER OF SEQUENCES: 26
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
;     STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
;     CITY: LOS ANGELES
;     STATE: CALIFORNIA
;     COUNTRY: U.S.A.
;     ZIP: 90067
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/405,234
;     FILING DATE: 13-MAR-1995
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: HAILE, Ph.D., LISA A.
;     REGISTRATION NUMBER: 38,347
;     REFERENCE/DOCKET NUMBER: PD3535
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 619/455-5100
;     TELEFAX: 619/455-5110
; INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 110 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 2..110
US-08-405-234-9

Alignment Scores:
 Pred. No.:              0.00721    Length:           110
 Score:                  61.00      Matches:           14
 Percent Similarity:     84.21%     Conservative:       2
 Best Local Similarity:  73.68%     Mismatches:         3
 Query Match:            67.78%     Indels:             0
 DB:                     1          Gaps:               0

US-10-045-180A-4 (1-19) x US-08-405-234-9 (1-110)

Qy  1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        ||||||||||||     :::||||::|||  ||||||||||||||||||| |||
Db  5 ATGAGGACCCTGCGCCATCCTGCTGCTGGTGGCCCTGCAGGCCCAGGCT 61

RESULT 11
US-08-158-189-10
; Sequence 10, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
;   APPLICANT: Bevins, Charles L.
;   APPLICANT: Jones, Douglas E.
```

; TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
;   STREET: One Liberty Place - 46th Floor
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: U.S.A.
;   ZIP: 19103
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/158,189
;   FILING DATE:
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/07/888,232
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Johnson, Philip S.
;   REGISTRATION NUMBER: 27,200
;   REFERENCE/DOCKET NUMBER: CH-0219
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-568-3100
;   TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 203 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: unknown
US-08-158-189-10

Alignment Scores:
Pred. No.:   0.015   Length: 203
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 67.78%   Indels: 0
DB: 1   Gaps: 0

US-10-045-180A-4 (1-19) x US-08-158-189-10 (1-203)

```
Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        |||||||||| ::|||||::||| |||||||||||||||||||||||||||||
Db  75 ATGAGGACCCTGCCATCCTTGCTGCCATTCTTCCTGGTGGCCCTGCGCAGGCCCAGGCT 131
```

RESULT 12
US-08-158-189-11
; Sequence 11, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
;   APPLICANT: Bevins, Charles L.
;   APPLICANT: Jones, Douglas E.
;   TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
;   TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
;   NUMBER OF SEQUENCES: 51
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
;   STREET: One Liberty Place - 46th Floor
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: U.S.A.
;   ZIP: 19103
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/158,189
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/07/888,232
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Johnson, Philip S.
;   REGISTRATION NUMBER: 27,200
;   REFERENCE/DOCKET NUMBER: CH-0219
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-568-3100
;   TELEFAX: 215-568-3439
;   INFORMATION FOR SEQ ID NO: 11:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 203 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: unknown
US-08-158-189-11

Alignment Scores:
Pred. No.:   0.015   Length: 203
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 67.78%   Indels: 0
DB: 1   Gaps: 0

US-10-045-180A-4 (1-19) x US-08-158-189-11 (1-203)

```
Qy   1 MetArgThrLeuThrLeuLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla 19
        |||||||||| ::|||||::||| |||||||||||||||||||||||||||||
Db  75 ATGAGGACCCTGCCATCCTTGCTGCCATTCTTCCTGGTGGCCCTGCGCAGGCCCAGGCT 131
```

RESULT 13
US-08-158-189-38
; Sequence 38, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
;   APPLICANT: Bevins, Charles L.
;   APPLICANT: Jones, Douglas E.
;   TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
;   TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
;   NUMBER OF SEQUENCES: 51
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
;   STREET: One Liberty Place - 46th Floor
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: U.S.A.
;   ZIP: 19103
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/158,189
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/07/888,232
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Johnson, Philip S.
;   REGISTRATION NUMBER: 27,200
;   REFERENCE/DOCKET NUMBER: CH-0219
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-568-3100
;   TELEFAX: 215-568-3439
;   INFORMATION FOR SEQ ID NO: 38:

Query Match: 65.56%   Indels: 0
DB: 1   Gaps: 0

US-10-045-180A-4 (1-19) x US-08-158-189-4 (1-424)

```
Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla  19
         ||||||||||::::  ::::||||::|||   ||||||||||||||||||||
Db  10 ATGAGGACCATCGCCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCGACCCCAGGCT  66
```

RESULT 15
US-08-021-608D-5/c
; Sequence 5, Application US/08021608D
; Patent No. 5580760
; GENERAL INFORMATION:
;   APPLICANT: LEVENS, DAVID L., DUNCAN,
;   APPLICANT: ROBERT C., AND AVIGAN, MARK I.
;   TITLE OF INVENTION: NOVEL FUSE BINDING
;   TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
;   NUMBER OF SEQUENCES: 24
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: MORGAN & FINNEGAN
;   STREET: 345 PARK AVENUE
;   CITY: NEW YORK
;   STATE: NEW YORK
;   COUNTRY: USA
;   ZIP: 10154
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: FLOPPY DISK
;   COMPUTER: IBM PC COMPATIBLE
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WORDPERFECT 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/021,608D
;   FILING DATE: 22-FEB-1993
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: WILLIAM S. FEILER
;   REGISTRATION NUMBER: 26,728
;   REFERENCE/DOCKET NUMBER: 2026-4063
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 758-4800
;   TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1097
;   TYPE: Nucleic acid
;   STRANDEDNESS: Double
;   TOPOLOGY: Unknown
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: No
;   ANTI-SENSE: YES
;   ORIGINAL SOURCE:
;   ORGANISM: Human
;   CELL LINE: HL60
US-08-021-608D-5

Alignment Scores:
Pred. No.:   0.252   Length: 1097
Score: 59.00   Matches: 13
Percent Similarity: 88.24%   Conservative: 2
Best Local Similarity: 76.47%   Mismatches: 2
Query Match: 65.56%   Indels: 0
DB: 1   Gaps: 0

US-10-045-180A-4 (1-19) x US-08-021-608D-5 (1-1097)

```
Qy    1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaAla  17
          |||||||||:::|||::::|||   ||||||||||||||||||||
Db 1086 ATGAGGACCCTCGCCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCC 1036
```

;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 218 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
US-08-158-189-38

Alignment Scores:
Pred. No.:   0.0163   Length: 218
Score: 61.00   Matches: 14
Percent Similarity: 84.21%   Conservative: 2
Best Local Similarity: 73.68%   Mismatches: 3
Query Match: 67.78%   Indels: 0
DB: Gaps: 0

US-10-045-180A-4 (1-19) x US-08-158-189-38 (1-218)

```
Qy   1 MetArgThrLeuThrLeuSerAlaPheLeuLeuValAlaLeuGlnAlaTrpAla  19
         ||||||||||  :::|||::::|||   ||||||||||||||||||||
Db  34 ATGAGGACCCTCGCCATCCTTGCTGCCATTCTCCTGGTGGCCCTGCAGGCCCCAGGCT  90
```

RESULT 14
US-08-158-189-4
; Sequence 4, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
;   APPLICANT: Bevins, Charles L.
;   APPLICANT: Jones, Douglas E.
;   TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
;   TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
;   NUMBER OF SEQUENCES: 51
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
;   STREET: One Liberty Place - 46th Floor
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: U.S.A.
;   ZIP: 19103
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/158,189
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/07/888,232
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Johnson, Philip S.
;   REGISTRATION NUMBER: 27,200
;   REFERENCE/DOCKET NUMBER: CH-0219
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-568-3100
;   TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 424 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 10..294
US-08-158-189-4

Alignment Scores:
Pred. No.:   0.0808   Length: 424
Score: 59.00   Matches: 13
Percent Similarity: 84.21%   Conservative: 3
Best Local Similarity: 68.42%   Mismatches: 3

Search completed: December 17, 2003, 14:54:00
Job time : 20 secs

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      December 17, 2003, 14:50:33 ; Search time 1258.91 Seconds
                                (without alignments)
                                1429.821 Million cell updates/sec

Title:         US-10-045-180A-5
Perfect score: 233
Sequence:      1 EPLQARAHEMPAQKQPPADD.........IYFSGDDSCSLQVPGSTKGL 44

Scoring table: BLOSUM62
               Xgapop  10.0 , Xgapext  0.5
               Ygapop  10.0 , Ygapext  0.5
               Fgapop   6.0 , Fgapext  7.0
               Delop    6.0 , Delext   7.0

Searched:     2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters:       5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145023_12528/app query.fasta_1.860
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10045180 @CGN_1_1_5283 @runat_17122003_145023_12528 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=30 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :    GenEmbl:*
              1:  gb_ba:*
              2:  gb_htg:*
              3:  gb_in:*
              4:  gb_om:*
              5:  gb_ov:*
              6:  gb_pat:*
              7:  gb_ph:*
              8:  gb_pl:*
              9:  gb_pr:*
             10:  gb_ro:*
             11:  gb_sts:*
             12:  gb_sy:*
             13:  gb_un:*
             14:  gb_vi:*
             15:  em_ba:*
             16:  em_fun:*
             17:  em_hum:*
             18:  em_in:*
             19:  em_mu:*
             20:  em_om:*
             21:  em_or:*
             22:  em_ov:*
             23:  em_pat:*
             24:  em_ph:*
             25:  em_pl:*
             26:  em_ro:*
             27:  em_sts:*
             28:  em_un:*
             29:  em_vi:*
             30:  em_htg_hum:*
             31:  em_htg_inv:*
             32:  em_htg_other:*
             33:  em_htg_mus:*
             34:  em_htg_pln:*
             35:  em_htg_rod:*
             36:  em_htg_mam:*
             37:  em_htg_vrt:*
             38:  em_sy:*
             39:  em_htgo_hum:*
             40:  em_htgo_mus:*
             41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		233	100.0	453	6	A98571	A98571 Sequence 2
2		233	100.0	453	6	BD074746	BD074746 Human def
3		209	89.7	4415	6	A98570	A98570 Sequence 1
4		209	89.7	4415	6	BD074745	BD074745 Human def
5	C	201	86.3	128544	9	AF238378	AF238378 Homo sapi
6		181	77.7	163093	2	AC116559	AC116559 Papio anu
7	C	181	77.7	188604	2	AC116558	AC116558 Papio anu
8		114	48.9	542	9	HSCCSHP4	X65977 H.sapiens m
9		108	46.4	4580	9	AF195487	AF195487 Homo sapi
10	C	108	46.4	79730	9	AF314060	AF314060 Homo sapi
11		108	46.4	150940	9	AF233439	AF233439 Homo sapi
12	C	108	46.4	158653	9	AF200455	AF200455 Homo sapi
13	C	108	46.4	168702	9	AC079018	AC079018 Homo sapi
14	C	108	46.4	169044	9	AF205406	AF205406 Homo sapi
15		104	44.6	496	9	AF355799	AF355799 Homo sapi
16		101	43.3	452	6	AR270635	AR270635 Sequence 6
17		101	43.3	452	6	I49588	I49588 Sequence 6
18		101	43.3	452	9	HUMDFSNSIX	M98331 Homo sapien
19		101	43.3	38845	9	AF314059	AF314059 Homo sapi
20		98	42.1	3208	9	HSU10267	U10267 Human defen
21		95	40.8	3377	9	AF104257	AF104257 Homo sapi
22		95	40.8	5124	9	HUNPIV	U18745 Human corti
23		90	38.6	398	6	AX405993	AX405993 Sequence
24		90	38.6	3060	9	HSU33317	U33317 Human defen
25	C	90	38.6	232182	2	AC021883	AC021883 Homo sapi
26		86	36.9	243	6	AR181803	AR181803 Sequence
27		86	36.9	421	9	AX047114	AX047114 Sequence
28		86	36.9	469	6	AF184157	AF184157 Macaca mu
29		84.5	36.3	469	4	RABDNP3AB	RABDNP3AB Oryctolagus
30		83	35.6	460	6	AF184156	AF184156 Macaca mu
31		83	35.6	495	6	AR181796	AR181796 Macaca mu
32		83	35.6	495	6	AX047101	AX047101 Sequence
33		83	35.6	495	6	AF191101	AF191101 Macaca mu
34		83	35.6	3090	4	RABDNP3AA	RABDNP3AA Oryctolagus
35		82	35.2	421	9	AF184158	AF184158 Macaca mu
36		82	35.2	500	6	AR181795	AR181795 Sequence
37		82	35.2	500	6	AX047099	AX047099 Sequence
38		82	35.2	500	6	AF191100	AF191100 Macaca mu
39		80	34.3	424	6	I49587	I49587 Sequence 4
40		80	34.3	2548	6	AX047111	AX047111 Sequence 4
41		80	34.3	2548	6	AF191103	AF191103 Macaca mu
42		80	34.3	2553	6	AR181800	AR181800 Sequence
43		78.5	33.7	448	9	HUMDEFIA	HUMDEFIA Human neutr
44		78.5	33.7	451	9	HUMDEFB	HUMDEFB Human defen
45		78.5	33.7	464	9	HSDEF3	X13621 Human HNP-3

ALIGNMENTS

RESULT 1

```
A98571
LOCUS       A98571                   453 bp    DNA     linear   PAT 26-JAN-2000
DEFINITION  Sequence 2 from Patent WO9911663.
ACCESSION   A98571
VERSION     A98571.1  GI:6781627
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Bougueleret,L. and Chumakov,I.
  TITLE     HUMAN DEFENSIN DEF-X, GENE AND DNAC, COMPOSITION CONTAINING SAME
            AND DIAGNOSTIC AND THERAPEUTIC APPLICATIONS
  JOURNAL   Patent: WO 9911663-A 2 11-MAR-1999;
            BOUGUELERET LYDIE (FR); CHUMAKOV ILYA (FR)
FEATURES             Location/Qualifiers
     source          1..453
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT    107 a    121 c     97 g    128 t
ORIGIN

Alignment Scores:
Pred. No.:               1.7e-24      Length:              453
Score:                    233.00      Matches:              44
Percent Similarity:      100.00%      Conservative:          0
Best Local Similarity:   100.00%      Mismatches:            0
Query Match:             100.00%      Indels:                0
DB:                            6      Gaps:                  0

US-10-045-180A-5 (1-44) x A98571 (1-453)

Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109 GAGCCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGAC 168

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  169 CAGGATGTGGTCATTTACTTTCAGGAGGATGACAGCTGCTCTCTTCAGGTTCCAGGCTCA 228

Qy   41 ThrLysGlyLeu 44
        ||||||||||||
Db  229 ACAAAGGGCTTG 240

RESULT 2
BD074746
LOCUS       BD074746                 453 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Human defensin polypeptide Def-X, genome DNA and cDNA, composition
            containing the same, and application to diagnosis and remedy
            treatment.
ACCESSION   BD074746
VERSION     BD074746.1  GI:22620349
KEYWORDS    JP 2001514264-A/2.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Theria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Bougueleret,L. and Shmacov,I.
  TITLE     Human defensin polypeptide Def-X, genome DNA and cDNA, composition
            containing the same, and application to diagnosis and remedy
            treatment.
  JOURNAL   Patent: JP 200151264-A 2 11-SEP-2001;
            GENSET
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001514264-A/2
            PD   11-SEP-2001
            PF   28-AUG-1998 JP 2000508701
            PR   29-AUG-1997 FR   97/10823
            PI   LYDIE BOUGUELERET,ILYA SHMACOV
            PC   C07K14/435,A01N43/50,A01N63/00,A61K7/00,A61K38/00,A61P29/00,
            PC   A61P35/00,
            PC   A61P37/02,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N15/09, PC
```

```
            C12P21/02,
            PC   C12P21/09,C12Q1/68,G01N33/53,A61K37/02,C12N15/00 CC
            CC   Strandedness: Double;
            CC   Topology: Linear;
            CC   Human defensin polypeptide Def-X, genome DNA and cDNA, CC
            CC   composition
            CC   containing the same, and application to diagnosis and remedy
                 treatment
            FH   Key             Location/Qualifiers
            FT   source          1..453
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1..453
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT    107 a    121 c     97 g    128 t
ORIGIN

Alignment Scores:
Pred. No.:               1.7e-24      Length:              453
Score:                    233.00      Matches:              44
Percent Similarity:      100.00%      Conservative:          0
Best Local Similarity:   100.00%      Mismatches:            0
Query Match:             100.00%      Indels:                0
DB:                            6      Gaps:                  0

US-10-045-180A-5 (1-44) x BD074746 (1-453)

Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109 GAGCCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGAC 168

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  169 CAGGATGTGGTCATTTACTTTCAGGAGGATGACAGCTGCTCTCTTCAGGTTCCAGGCTCA 228

Qy   41 ThrLysGlyLeu 44
        ||||||||||||
Db  229 ACAAAGGGCTTG 240

RESULT 3
A98570
LOCUS       A98570                  4415 bp    DNA     linear   PAT 26-JAN-2000
DEFINITION  Sequence 1 from Patent WO9911663.
ACCESSION   A98570
VERSION     A98570.1  GI:6781626
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4415)
  AUTHORS   Bougueleret,L. and Chumakov,I.
  TITLE     HUMAN DEFENSIN DEF-X, GENE AND DNAC, COMPOSITION CONTAINING SAME
            AND DIAGNOSTIC AND THERAPEUTIC APPLICATIONS
  JOURNAL   Patent: WO 9911663-A 1 11-MAR-1999;
            BOUGUELERET LYDIE (FR); CHUMAKOV ILYA (FR)
FEATURES             Location/Qualifiers
     source          1..4415
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
     exon            1836..1874
     exon            3394..3577
     exon            4161..4380
     polyA_site      4374..4379
BASE COUNT   1128 a   1109 c    912 g   1252 t
ORIGIN

Alignment Scores:
Pred. No.:              6.65e-20      Length:             4415
Score:                    209.00      Matches:              39
Percent Similarity:      100.00%      Conservative:          0
```

Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 89.70%   Indels: 0
DB: 6   Gaps: 0

US-10-045-180A-5 (1-44) x A98570 (1-4415)

Qy   1 GluProLeuGlnAlaAArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3463 GAGCCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGAC 3522

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly 39
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3523 CAGGATGTGGTCATTTACTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGT 3579

RESULT 5
AF238378/c
LOCUS   AF238378   128544 bp   DNA   linear   PRI 02-APR-2003
DEFINITION   Homo sapiens chromosome 8 clone SCb-561b17 map p22-p21, complete
              sequence.
ACCESSION   AF238378
VERSION   AF238378.5   GI:29469504
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 128544)
  AUTHORS   Polley,A., Baumgart,C., Blechschmidt,K., Dette,M.D., Jahn,N.,
              Menzel,U., Reichwald,K., Schilhabel,M.B., Schudy,A., Taudien,S.,
              Wen,G., Schutte,B., Malik,M., Peng,J.Hong., McCray,P. and
              Rosenthal,A.
  TITLE   Chromosome 8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 128544)
  AUTHORS   Schudy,A., Schilhabel,M., Schutte,B., Ganz,T., Linzmeier,R.,
              Ho,C.H., Hoang,B.V., McCray,P., Baumgart,C., Menzel,U.,
              Schattevoy,R. and Rosenthal,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (22-FEB-2000) Genome Analysis, Institute of Molecular
              Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   3 (bases 1 to 128544)
  AUTHORS   Genome Sequencing Center Jena.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular
              Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4 (bases 1 to 128544)
  AUTHORS   Genome Sequencing Center Jena.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-NOV-2000) Genome Analysis, Institute of Molecular
              Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5 (bases 1 to 128544)
  AUTHORS   Lagemann,D. and Platzer,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (06-JUL-2002) Genome Analysis, Institute of Molecular
              Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   6 (bases 1 to 128544)
  AUTHORS   Lagemann,D. and Platzer,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-APR-2003) Genome Analysis, Institute of Molecular
              Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT   On Apr 2, 2003 this sequence version replaced gi:21700555.
              --------------------- Genome Center
                 Center: Insitute of Molecular Biotechnology
                 Center code: IMB
                 Web site: http://genome.imb-jena.de/
                 Contact: gscj-submit@genome.imb-jena.de
              --------------------- Project Information
                 Center project name: H370
                 Center clone name: SCb-561b17
              --------------------- Summary Statistics
                 Sequencing vector: pUC18; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990329
                 Consensus quality: 127999 bases at least Q40
                 Consensus quality: 128330 bases at least Q30
                 Consensus quality: 128544 bases at least Q20
                 Quality coverage: 11.05x

Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 89.70%   Indels: 0
DB: 6   Gaps: 0

US-10-045-180A-5 (1-44) x A98570 (1-4415)

Qy   1 GluProLeuGlnAlaAArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3463 GAGCCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGAC 3522

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly 39
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3523 CAGGATGTGGTCATTTACTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGT 3579

RESULT 4
BD074745
LOCUS   BD074745   4415 bp   DNA   linear   PAT 27-AUG-2002
DEFINITION   Human defensin polypeptide Def-X, genome DNA and cDNA, composition
              containing the same, and application to diagnosis and remedy
              treatment.
ACCESSION   BD074745
VERSION   BD074745.1   GI:22620348
KEYWORDS   JP 2001514264-A/1.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 4415)
  AUTHORS   Bougueleret,L. and Shmacov,I.
  TITLE   Human defensin polypeptide Def-X, genome DNA and cDNA, composition
              containing the same, and application to diagnosis and remedy
  JOURNAL   Patent: JP 2001514264-A 1 11-SEP-2001;
              GENSET
COMMENT   OS   Homo sapiens (human)
              PN   JP 2001514264-A/1
              PD   11-SEP-2001
              PF   28-AUG-1998 JP 2000508701
              PR   29-AUG-1997 FR   97/10823
              PI   LYDIE BOUGUELERET,ILYA SHMACOV
              PC   C07K14/435,A01N43/50,A01N63/00,A61K7/00,A61K38/00,A61K29/00,
              PC   A61P35/00,
              PC   A61P37/02,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N15/09, PC
              C12N1/02,
              PC   C12P21/08,C12Q1/68,G01N33/53,A61K37/02,C12N15/00 CC
              Strandedness: Double;
              CC   Topology: Linear;
              CC   Human defensin polypeptide Def-X, genome DNA and cDNA,   CC
              composition
              CC   containing the same, and application to diagnosis and remedy
              CC   treatment
FEATURES   Location/Qualifiers
     source   1..4415
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     FH   Key   Location/Qualifiers
     FT   exon   1836..1874
     FT   exon   3394..3577
     FT   exon   4161..4380
     FT   CDS   3406..3408
     FT   CDS   4276..4278
BASE COUNT   1128 a   1109 c   912 g   1252 t   14 others
ORIGIN

Alignment Scores:
Pred. No.:   6.65e-20   Length: 4415
Score: 209.00   Matches: 39
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 89.70%   Indels: 0
DB: 6   Gaps: 0

US-10-045-180A-5 (1-44) x BD074745 (1-4415)

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

```
FEATURES             Location/Qualifiers
     source          1..128544
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="p22-p21"
                     /clone="SCb-561b17"
     misc_feature    44529..44595
                     /note="single stranded/single chemistry region"
     misc_feature    45703
                     /note="low quality region"
     misc_feature    45815..45872
                     /note="low quality region"
     misc_feature    45824..45872
                     /note="single stranded/single chemistry region"
     unsure          45843
     unsure          45847
     unsure          45849
     unsure          45858
     unsure          45868
     misc_feature    46235
                     /note="low quality region"
     misc_feature    47899..47900
                     /note="low quality region"
     unsure          47899..47900
     misc_feature    49172
                     /note="low quality region"
     unsure          49172
     misc_feature    49220
                     /note="low quality region"
     misc_feature    49339
                     /note="low quality region"
     misc_feature    49756..49807
                     /note="low quality region"
     misc_feature    49961
                     /note="single stranded/single chemistry region"
     misc_feature    52933..52948
                     /note="low quality region"
     unsure          52943
     misc_feature    53544
                     /note="low quality region"
     misc_feature    53612..53926
                     /note="low quality region"
     misc_feature    53674..53677
                     /note="single stranded/single chemistry region"
     misc_feature    53779..53799
                     /note="low quality region"
     unsure          53787
     misc_feature    54734..55129
                     /note="single stranded/single chemistry region"
     misc_feature    56028..56165
                     /note="single stranded/single chemistry region"
     misc_feature    57952..58060
                     /note="single stranded/single chemistry region"

     misc_feature    57968..57974
                     /note="low quality region"
     unsure          57968
     unsure          57970
     unsure          57974
     misc_feature    58652
                     /note="low quality region"
     misc_feature    59561..59638
                     /note="single stranded/single chemistry region"
     misc_feature    59925
                     /note="low quality region"
     unsure          59925
     misc_feature    59929
                     /note="low quality region"
     misc_feature    61695..62103
                     /note="single stranded/single chemistry region"
     misc_feature    61858..61859
                     /note="low quality region"
     misc_feature    62263..62485
                     /note="single stranded/single chemistry region"
     misc_feature    62655..63118
                     /note="single stranded/single chemistry region"
     misc_feature    62750..62785
                     /note="low quality region"
     unsure          62774..62777
     unsure          62783
     misc_feature    62801..62825
                     /note="low quality region"
     unsure          62819
     unsure          62822
     misc_feature    62872..62874
                     /note="low quality region"
     misc_feature    62908
                     /note="low quality region"
     misc_feature    62917..62987
                     /note="low quality region"
     unsure          62933..62935
     unsure          62955
     unsure          62957
     misc_feature    63082
                     /note="low quality region"
     misc_feature    63110
                     /note="low quality region"
     misc_feature    63119..63281
                     /note="low quality region"
     misc_feature    63119..63141
                     /note="single clone coverage"
     unsure          63129..63131
     misc_feature    63142..63258
                     /note="single stranded/single chemistry region"
     unsure          63159
     unsure          63197..63258
     misc_feature    63281
                     /note="low quality region"
     misc_feature    63876..63878
                     /note="low quality region"
     misc_feature    63933..64021
                     /note="single stranded/single chemistry region"
     misc_feature    63943
                     /note="low quality region"
     misc_feature    63984..63985
                     /note="low quality region"
     misc_feature    64004..64007
                     /note="low quality region"
     misc_feature    70360
                     /note="low quality region"
```

Alignment Scores:
Pred. No.:   4.33e-17   Length: 128544
Score: 201.00   Matches: 37
Percent Similarity: 97.44%   Conservative: 1
Best Local Similarity: 94.87%   Mismatches: 1
Query Match: 86.27%   Indels: 0
DB: 9   Gaps: 0

Percent Similarity:　94.87%　　　Conservative:　2
Best Local Similarity:　89.74%　　Mismatches:　2
Query Match:　77.68%　　　Indels:　0
DB:　2　　　Gaps:　0

US-10-045-180A-5 (1-44) x AC16559 (1-163093)

Qy　1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
|||||||||||||||||||||||||||||||||||||||||||:::|||
Db　32923 GAGCCGCTCCAGGCAAGAGCTGATGAGATGCCAGCCCAGGAGCAGCCTCAGCAAACGAC 32982

Qy　21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly 39
||||||||||||||||||||||||||:|||||||||||||||||||
Db　32983 CAAGATGTGGTCATTTACTTTCAGGAGGATGACAGCGTCCTTCTTCAGGTTCCAGGT 33039

RESULT 7
AC116558/c
LOCUS　　AC116558　　　188604 bp　　DNA　　linear　HTG 11-JUN-2003
DEFINITION　Papio anubis clone rp41-273g19, WORKING DRAFT SEQUENCE, 2 ordered
　　　　　pieces.
ACCESSION　AC116558
VERSION　AC116558.16　GI:30725961
KEYWORDS　HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE　　Papio anubis (olive baboon)
ORGANISM　Papio anubis
　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
　　　　　Cercopithecinae; Papio.
REFERENCE　1 (bases 1 to 188604)
AUTHORS　Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE　　Papio anubis BAC Clone rp41-273g19
JOURNAL　Unpublished
REFERENCE　2 (bases 1 to 188604)
AUTHORS　Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE　　Direct Submission
JOURNAL　Submitted (29-MAR-2002) Department Of Chemistry And Biochemistry,
　　　　　The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
　　　　　OK 73019, USA
REFERENCE　3 (bases 1 to 188604)
AUTHORS　Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE　　Direct Submission
JOURNAL　Submitted (11-JUN-2003) Department Of Chemistry And Biochemistry,
　　　　　The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
　　　　　OK 73019, USA
COMMENT　On May 15, 2003 this sequence version replaced gi:28173120.
　　　　　----------- Genome Center
　　　　　Center: Department Of Chemistry And Biochemistry
　　　　　The University of Oklahoma
　　　　　Center code:UOKNOR
　　　　　-----------
　　　　* NOTE: This is a 'working draft' sequence. It currently
　　　　* consists of 2 contigs. Gaps between the contigs
　　　　* are represented as runs of N. The order of the pieces
　　　　* is believed to be correct as given, however the sizes
　　　　* of the gaps between them are based on estimates that have
　　　　* provided by the submittor.
　　　　* This sequence will be replaced
　　　　* by the finished sequence as soon as it is available and
　　　　* the accession number will be preserved.
　　　　*　　1　　6057: contig of 6057 bp in length
　　　　*　6058　6157: gap of unknown length
　　　　*　6158　188604: contig of 182447 bp in length.
FEATURES　Location/Qualifiers
　source　　1..188604
　　　　　/organism="Papio anubis"
　　　　　/mol_type="genomic DNA"
　　　　　/db_xref="taxon:9555"
　　　　　/clone=rp41-273g19"
　　　　　/clone_lib="RPCI - 41 Male (Olive) Baboon BAC Library"
BASE COUNT　53384 a　41769 c　40796 g　52555 t　100 others
ORIGIN
Alignment Scores:

US-10-045-180A-5 (1-44) x AF238378 (1-128544)

Qy　1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
|||||||||||||||||||||||||||||||||||||||||||||||
Db　6266 GAGCCGCTCCGGCAAGAGCTCATGAGATGCCAGCCCAAGCAGCCTCCAGCAGATGAC 63207

Qy　21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly 39
|||||||||:::|||||||||||||||||||||||||||||||||
Db　6306 CAGGATGTGGTCCTTTACTTTCAGGAGGATGACAGCGTCTTCTTCAGTTCCAGGT 63150

RESULT 6
AC116559
LOCUS　　AC116559　　　163093 bp　　DNA　　linear　HTG 30-MAY-2003
DEFINITION　Papio anubis clone rp41-339c10, WORKING DRAFT SEQUENCE, 3 ordered
　　　　　pieces.
ACCESSION　AC116559
VERSION　AC116559.18　GI:31193968
KEYWORDS　HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE　　Papio anubis (olive baboon)
ORGANISM　Papio anubis
　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
　　　　　Cercopithecinae; Papio.
REFERENCE　1 (bases 1 to 163093)
AUTHORS　Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE　　Papio anubis BAC Clone rp41-339c10
JOURNAL　Unpublished
REFERENCE　2 (bases 1 to 163093)
AUTHORS　Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE　　Direct Submission
JOURNAL　Submitted (29-MAR-2002) Department Of Chemistry And Biochemistry,
　　　　　The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
　　　　　OK 73019, USA
REFERENCE　3 (bases 1 to 163093)
AUTHORS　Xu,W., Hua,A., Eichler,E. and Roe,B.A.
TITLE　　Direct Submission
JOURNAL　Submitted (30-MAY-2003) Department Of Chemistry And Biochemistry,
　　　　　The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
　　　　　OK 73019, USA
COMMENT　On May 30, 2003 this sequence version replaced gi:30725962.
　　　　　----------- Genome Center
　　　　　Center: Department Of Chemistry And Biochemistry
　　　　　The University Of Oklahoma
　　　　　Center code:UOKNOR
　　　　　-----------
　　　　* NOTE: This is a 'working draft' sequence. It currently
　　　　* consists of 3 contigs. Gaps between the contigs
　　　　* are represented as runs of N. The order of the pieces
　　　　* is believed to be correct as given, however the sizes
　　　　* of the gaps between them are based on estimates that have
　　　　* provided by the submittor.
　　　　* This sequence will be replaced
　　　　* by the finished sequence as soon as it is available and
　　　　* the accession number will be preserved.
　　　　*　　1　　6315: contig of 6315 bp in length
　　　　*　6316　6415: gap of unknown length
　　　　*　6416　16704: contig of 10289 bp in length
　　　　*　16705　16804: gap of unknown length
　　　　*　16805　163093: contig of 146289 bp in length.
FEATURES　Location/Qualifiers
　source　　1..163093
　　　　　/organism="Papio anubis"
　　　　　/mol_type="genomic DNA"
　　　　　/db_xref="taxon:9555"
　　　　　/clone_lib="rp41-339c10"
　　　　　/clone_lib="RPCI - 41 Male (Olive) Baboon BAC Library"
BASE COUNT　44435 a　36508 c　35861 g　46065 t　224 others
ORIGIN

Alignment Scores:
Pred. No.:　　　4.44e-14　　　Length:　　　163093
Score:　　　　181.00　　　Matches:　　　35

```
Pred. No.:                    5.23e-14        Length:              188604
Score:                        181.00          Matches:             35
Percent Similarity:           94.87%          Conservative:        2
Best Local Similarity:        89.74%          Mismatches:          2
Query Match:                  77.68%          Indels:              0
DB:                                           Gaps:                0

US-10-045-180A-5 (1-44) x AC116558 (1-188604)

Qy         1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
             |||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db    142609 GAGCCGCTCCAGGCAAGAGCTGATGAGATGCCAGCCCAGGAGCAGCCTCCAGCAAACGAC 142550

Qy        21 GlnAspValValIleTyrPheSerGlyAspSerCysSerLeuGlnValProGly 39
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142549 CAAGATGTGGTCATTTACTTTTCAGGAGATGACAGCTCCTCTTCAGGTTCCAGGT 142493

RESULT 8
HSCCSHP4
LOCUS       HSCCSHP4                  542 bp    mRNA    linear    PRI 07-JUN-1999
DEFINITION  H.sapiens mRNA for corticostatin HP-4 precursor.
ACCESSION   X65977
VERSION     X65977.1  GI:29734
KEYWORDS    corticostatin/defensin.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 542)
AUTHORS     Palfree,R.G., Sadro,L.C. and Solomon,S.
TITLE       The gene encoding the human corticostatin HP-4 precursor contains a
            recent 86-base duplication and is located on chromosome 8
JOURNAL     Mol. Endocrinol. 7 (2), 199-205 (1993)
MEDLINE     93225961
PUBMED      8469233
REFERENCE   2  (bases 1 to 542)
AUTHORS     Palfree,R.G.E.
TITLE       Direct Submission
JOURNAL     Submitted (05-MAY-1992) R.G.E. Palfree, McGill University Royal
            Victoria Hosp., Rm L2.05, 687 Pine Ave West, Montreal PQ, H3A 1A1,
            CANADA
FEATURES            Location/Qualifiers
     source         1..542
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /chromosome="8"
                    /clone="HP-4.1"
                    /tissue_type="bone marrow"
                    /clone_lib="Clontech HL1058a, human bone marrow cDNA"
                    /dev_stage="adult"
     gene           1..542
                    /gene="HP-4"
     CDS            52..345
                    /gene="HP-4"
                    /codon_start=1
                    /product="corticostatin/defensin HP-4 precursor"
                    /protein_id="CAA46792.1"
                    /db_xref="GI:29735"
                    /db_xref="SWISS-PROT:P12838"
                    /translation="MRIIALLAAILLVALQVRAGPIQARGDEAPGQBQRGPEDQDISI
                    SFAWDKSSALQVSGSTRGMVCSCRLVFCRRTELRVGNCLIGGVSFTYCCTRVD"
     sig_peptide    52..108
                    /gene="HP-4"
     mat_peptide    109..342
                    /gene="HP-4"
     misc_feature   330..412
                    /gene="HP-4"
                    /note="recent duplication of immediate upstream segment"
BASE COUNT      118 a     137 c     135 g     152 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                    3.46e-07        Length:              542
Score:                        114.00          Matches:             23
Percent Similarity:           69.77%          Conservative:        7
Best Local Similarity:        53.49%          Mismatches:          13
Query Match:                  48.93%          Indels:              0
DB:                                           Gaps:                0

US-10-045-180A-5 (1-44) x HSCCSHP4 (1-542)

Qy         2 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 21
             ||||||||||||||||||||||||||||||||  |||:::|||       ::::|||
Db       112 CCACTCCAGGCAAGAGCTGATGAGATGCCAGCCCAGGAGCAGCCGTGGGCCAGAAGACCAG 171

Qy        22 AspValValIleTyrPheSerGlyAspSerCysSerLeuGlnValProGlySerThr 41
             |||::|    ||||||||||||||||||||     :::|||||||||||||
Db       172 GACATATCTATTCCTTGCATGGGATAAAAGCTCTGCTCTTCAGGTCAACA 231

Qy        42 LysGlyLeu 44
             :::||||::
Db       232 AGGGGCATG 240

RESULT 9
AF195487
LOCUS       AF195487                 4580 bp   DNA     linear    PRI 09-DEC-1999
DEFINITION  Homo sapiens defensin-related pseudogene.
ACCESSION   AF195487
VERSION     AF195487.1  GI:6540006
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4580)
AUTHORS     Bevins,C.L. and Jones,D.E.
TITLE       Analysis of the Human Defensin Gene Locus: Sequence of Clone HG27B
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 4580)
AUTHORS     Bevins,C.L. and Jones,D.E.
TITLE       Direct Submission
JOURNAL     Submitted (15-OCT-1999) Immunology, The Cleveland Clinic
            Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, USA
FEATURES            Location/Qualifiers
     source         1..4580
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="8"
                    /map="8p23"
     gene           1..4580
                    /gene="defensin-related pseudogene"
                    /pseudo
BASE COUNT     1194 a    1175 c    950 g    1261 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                    2.86e-05        Length:              4580
Score:                        108.00          Matches:             24
Percent Similarity:           68.29%          Conservative:        4
Best Local Similarity:        58.54%          Mismatches:          13
Query Match:                  46.35%          Indels:              0
DB:                                           Gaps:                0

US-10-045-180A-5 (1-44) x AF195487 (1-4580)

Qy         1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
             |||||||||||| |||    ||| ||||||||||:::||||||||||||
Db      3282 GAGCCACTCCAGGCAATTGCTGATGAGGCTACAGCCCAGGAGCAGCCTGGAGCAGATGAT 3341

Qy        21 GlnAspValValIleTyrPheSerGlyAspSerCysSerLeuGlnValProGlySer 40
             |||::|    |||||||    ||||||
Db      3342 CAGGAAGTGGTTGATTCCTTGCATGGGATGAAGAGTCCTCTTCAGGTTCAGGTAAG 3401
```

```
              GenCore version 5.1.6
              Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:       December 17, 2003, 14:50:33 ; Search time 1538.36 Seconds
                                           (without alignments)
                                           695.153 Million cell updates/sec
```

```
Title:        US-10-045-180A-5
Perfect score: 233
Sequence:     1 EPLQARAHEMPAQKQPPADD..........IYFSGDDSCSLQVPGSTKGL 44

Scoring table: BLOSUM62
              Xgapop  10.0 , Xgapext   0.5
              Ygapop  10.0 , Ygapext   0.5
              Fgapop   6.0 , Fgapext   7.0
              Delop    6.0 , Delext    7.0

Searched:     22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters:       45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries
```

```
Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US1045180/runat_17122003_145024_12540/app_query.fasta_1.860
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1045180 @CGN_1_1_6100 @runat_17122003_145024_12540 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

```
Database :    EST:*
          1:  em_estba:*
          2:  em_esthum:*
          3:  em_estin:*
          4:  em_estmu:*
          5:  em_estov:*
          6:  em_estpl:*
          7:  em_estro:*
          8:  em_htc:*
          9:  gb_est1:*
         10:  gb_est2:*
         11:  gb_htc:*
         12:  gb_est3:*
         13:  gb_est4:*
         14:  gb_est5:*
         15:  em_estfun:*
         16:  em_estom:*
         17:  em_gss_hum:*
         18:  em_gss_inv:*
         19:  em_gss_pln:*
         20:  em_gss_vrt:*
         21:  em_gss_fun:*
         22:  em_gss_mam:*
         23:  em_gss_mus:*
         24:  em_gss_pro:*
         25:  em_gss_rod:*
         26:  em_gss_phg:*
         27:  em_gss_vrl:*
         28:  gb_gss1:*
```

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	114	48.9	307	12	BI021150	BI021150 CM0-MT035
c	2	114	48.9	422	13	BX099382	BX099382 BX099382
c	3	114	48.9	458	9	AI082774	AI082774 ox74a06.x
c	4	114	48.9	541	9	AA722853	AA722853 zg85h07.s
c	5	114	48.9	542	9	AI250799	AI250799 qi36g07.x
c	6	114	48.9	634	13	BU616655	BU616655 UI-H-DF0-
c	7	108	46.4	788	13	BX117347	BX117347 BX117347
c	8	107	45.9	682	29	AG171073	AG171073 Pan trog1
c	9	93.5	40.1	476	9	AI081728	AI081728 ow81f07.s
c	10	81.5	35.0	319	12	BI021560	BI021560 CM3-MT034
	11	80	34.3	249	12	BM830493	BM830493 K-EST0104
	12	80	34.3	300	12	BM765919	BM765919 K-EST0047
	13	80	34.3	303	12	BM855729	BM855729 K-EST0138
	14	80	34.3	351	12	BM768678	BM768678 K-EST0051
	15	80	34.3	375	12	BM856063	BM856063 K-EST0139
	16	80	34.3	376	12	BM855276	BM855276 K-EST0138
	17	80	34.3	443	12	BM855593	BM855593 K-EST0138
	18	80	34.3	447	12	BM703419	BM703419 UI-E-CL1-
	19	80	34.3	460	12	BM768516	BM768516 K-EST0051
	20	80	34.3	467	12	BM854660	BM854660 K-EST0137
	21	80	34.3	511	12	BM715953	BM715953 UI-E-BJ0-
c	22	80	34.3	513	12	BM674264	BM674264 UI-E-BJ0-
c	23	80	34.3	519	13	BU729350	BU729350 UI-E-CL1-
	24	78.5	33.7	208	14	R29450	R29450 F1-57I.D 22
	25	78.5	33.7	281	9	AA321281	AA321281 EST23811
	26	78.5	33.7	294	9	AA321316	AA321316 EST23783
	27	78.5	33.7	306	10	BF893430	BF893430 QV3-MT012
c	28	78.5	33.7	309	12	BI022856	BI022856 CM4-MT024
	29	78.5	33.7	312	9	AA321102	AA321102 EST23896
	30	78.5	33.7	320	9	AA321426	AA321426 EST23923
c	31	78.5	33.7	325	10	BF893434	BF893434 QV3-MT012
	32	78.5	33.7	327	9	AA321192	AA321192 EST23994
	33	78.5	33.7	330	10	BE831394	BE831394 QV4-MT005
	34	78.5	33.7	348	10	BF893385	BF893385 QV3-MT012
c	35	78.5	33.7	349	12	BI021617	BI021617 CM3-MT034
	36	78.5	33.7	354	9	AA321384	AA321384 EST23855
	37	78.5	33.7	355	10	BF893379	BF893379 QV3-MT012
	38	78.5	33.7	359	12	BI021220	BI021220 CM0-MT035
c	39	78.5	33.7	370	10	BF893376	BF893376 QV3-MT012
	40	78.5	33.7	370	10	BF893383	BF893383 QV3-MT012
	41	78.5	33.7	371	10	BF893218	BF893218 QV3-MT012
	42	78.5	33.7	372	10	BF893378	BF893378 QV3-MT012
	43	78.5	33.7	377	9	AA321224	AA321224 EST23726
	44	78.5	33.7	379	12	BI021556	BI021556 CM3-MT034
	45	78.5	33.7	379	14	C01361	C01361 HUMGS000834

```
                                                    ALIGNMENTS

RESULT 1
BI021150/c
LOCUS       BI021150     CM0-MT0351-150101-736-b03 MT0351 Homo sapiens cDNA,  307 bp     mRNA     linear     EST 14-JUN-2001
DEFINITION  CM0-MT0351-150101-736-b03 MT0351 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI021150
VERSION     BI021150.1  GI:14427780
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 307)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
```

Left column

```
          Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
          Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
  TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
  JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE 20202663
  PUBMED  10737800
COMMENT   Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM0&t2=CM0-MT0351-
          15101-736-b03&t3=2001-01-15&t4=1)
          Seq primer: puc 18 forward
          High quality sequence start: 21
          High quality sequence stop: 307.
FEATURES            Location/Qualifiers
     source         1..307
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /dev_stage="Adult"
                    /clone_lib="MT0351"
                    /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                    SmaI; A mini-library was made by cloning products derived
                    from ORESTES PCR (U.S. Letters Patent application No. 196
                    ,716 - Ludwig Institute for Cancer Research) profiles
                    into the pUC 18 vector. Reverse transcription of tissue
                    mRNA and cDNA amplification were performed under low
                    stringency conditions."
BASE COUNT      75 a     88 c     76 g     68 t
ORIGIN

Alignment Scores:
Pred. No.:                 1.06e-05          Length:            307
Score:                     114.00            Matches:           23
Percent Similarity:        69.77%            Conservative:      7
Best Local Similarity:     53.49%            Mismatches:        13
Query Match:               48.93%            Indels:            0
DB:                        12                Gaps:              0

US-10-045-180A-5 (1-44) x BI021150 (1-307)

Qy    2 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 21
        ||||||||||  ||  ||||:|||
Db  286 CCACTCCAGGCAAGAGGTGATGAGGCTCCAGGCAGACAGCGTGGGCCAGAAGACCAG 227

Qy   22 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 41
        |||:||:  |||   :::||||||||
Db  226 GACATATCTATTTCCTTTGCATGGGATAAAAGCTCTGCTCTTCAGGTTTCAGGCTCAACA 167

Qy   42 LysGlyLeu 44
        :::||||:::
Db  166 AGGGGCATG 158

RESULT 2
BX099382
LOCUS       BX099382      422 bp    mRNA    linear    EST 06-FEB-2003
DEFINITION  BX099382 Soares NhHMPu_S1 Homo sapiens cDNA clone IMAGp998H134549 ;
            IMAGE:1858620, mRNA sequence.
ACCESSION   BX099382
VERSION     BX099382.1  GI:27844054
KEYWORDS    EST.
  SOURCE    Homo sapiens (human)
```

Right column

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 422)
  AUTHORS   Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
            Radelof,U., Schneider,D. and Korn,B.
  TITLE     Human UnigeneSet - RZPD3
  JOURNAL   Unpublished
  COMMENT   Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
            RZPD; IMAGp998H134549.
            RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
            Human UnigeneSet - RZPD3 (R2RDLIB No.972)
            http://www.rzpd.de/CloneCards/cgi-
            bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 101
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available royalty-free from RZPD;
            contact RZPD (clone@rzpd.de) for further information. Seq primer:
            M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1..422
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGp998H134549 ; IMAGE:1858620"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /clone_lib="Soares NhHMPu_S1"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT      83 a     115 c     115 g     107 t     2 others
ORIGIN

Alignment Scores:
Pred. No.:                 1.58e-05          Length:            422
Score:                     114.00            Matches:           23
Percent Similarity:        69.77%            Conservative:      7
Best Local Similarity:     53.49%            Mismatches:        13
Query Match:               48.93%            Indels:            0
DB:                        13                Gaps:              0

US-10-045-180A-5 (1-44) x BX099382 (1-422)

Qy    2 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 21
        ||||||||||  ||  ||||:||:
Db  130 CCACTCCAGGCAAGAGGTGATGAGGCTCCAGGCAGACAGCGTGGGCCAGAAGACCAG 189

Qy   22 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 41
        |||:||:  |||   :::||||||||
Db  190 GACATATCTATTTCCTTTGCATGGGATAAAAGCTCTGCTCTTCAGGTTTCAGGCTCAACA 249

Qy   42 LysGlyLeu 44
        :::||||:::
Db  250 AGGGGCATG 258

RESULT 3
AI082774/c
LOCUS      AI082774      458 bp    mRNA    linear    EST 24-SEP-1998
```

DEFINITION   ox74a06.x1 Soares NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1662034
              3' similar to gb:X65977 NEUTROPHIL DEFENSIN 4 PRECURSOR (HUMAN);,
              mRNA sequence.
ACCESSION   AI082774
VERSION   AI082774.1 GI:3417750
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 458)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Insert Length: 637   Std Error: 0.00
              Seq primer: -40m13 fwd. ET from Amersham
              High quality sequence stop: 446.
FEATURES   Location/Qualifiers
     source   1..458
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:1662034"
              /tissue_type="Pooled human melanocyte, fetal heart, and
              pregnant uterus"
              /lab_host="DH10B"
              /clone_lib="Soares NhHMPu_S1"
              /note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I;
              Site 2: Eco RI; Equal amounts of plasmid DNA from three
              normalized libraries (melanocyte 2NbHM, pregnant uterus
              NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
              were made in vitro. Following HAP purification, this DNA
              was used as tracer in a subtractive hybridization
              reaction. The driver was PCR-amplified cDNAs from pools of
              5,000 clones made from the same 3 libraries. The pools
              consisted of I.M.A.G.E. clones 260232-265223,
              340488-345479, and 484488-489479."
BASE COUNT   128 a   117 c   103 g   109 t   1 others
ORIGIN

Alignment Scores:
Pred. No.:   1.74e-05
Score: 114.00
Percent Similarity: 69.77%
Best Local Similarity: 53.49%
Query Match: 48.93%
DB: 9

Length: 458
Matches: 23
Conservative: 7
Mismatches: 13
Indels: 0
Gaps: 0

US-10-045-180A-5 (1-44) x AI082774 (1-458)

Qy   2 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 21
          ||||||||| :::||||:||| ||::||| :::|||||||
Db   371 GACATATCTATTTCCTTTGCATGGGATAAAAGCTCTGCTTCTTCAGGTTCAGGCTCAACA 312

Qy   22 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 41
          |||:|||||||||||||||||| |||:::|||||||||||||||||
Db   431 CCACTCCGGCCAAGAGGTGATGAGGCTCCAGGCCAGGCGTGGGCCAGAAGACCAG 372

Qy   42 LysGlyLeu 44
          :::||:|||
Db   311 AGGGGCATG 303

RESULT 4
AA722853
LOCUS   AA722853   541 bp   mRNA   linear   EST 02-JAN-1998
DEFINITION zg85h07.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:409405 3' similar to gb:X65977 NEUTROPHIL DEFENSIN 4

PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   AA722853
VERSION   AA722853.1 GI:2740560
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 541)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
              Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
              ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
              White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE   WashU-NCI human EST Project
JOURNAL   Unpublished
COMMENT   Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Seq primer: -40m13 fwd. ET from Amersham
              High quality sequence stop: 484.
FEATURES   Location/Qualifiers
     source   1..541
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="GDB:1316190"
              /db_xref="taxon:9606"
              /clone="IMAGE:409405"
              /sex="unknown"
              /dev_stage="19 weeks"
              /lab_host="DH10B (ampicillin resistant)"
              /clone_lib="Soares_fetal_heart_NbHH19W"
              /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
              modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
              strand cDNA was primed with a Not I - oligo(dT) primer [5'
              TGTTACCAATCTGAAGTGGGAGCGGCGCGCCATCTTTTTTTTTTTTTTTTTTT 3'],
              double-stranded cDNA was size selected, ligated to Eco RI
              adapters (Pharmacia), digested with Not I and cloned into
              the Not I and Eco RI sites of a modified pT7T3 vector
              (Pharmacia). Library went through one round of
              normalization to a Cot = 5. Library constructed by
              M.Fatima Bonaldo. This library was constructed from the
              same fetus as the fetal lung library, Soares fetal lung
              NbHL19W."
BASE COUNT   152 a   133 c   137 g   119 t
ORIGIN

Alignment Scores:
Pred. No.:   2.15e-05
Score: 114.00
Percent Similarity: 69.77%
Best Local Similarity: 53.49%
Query Match: 48.93%
DB: 9

Length: 541
Matches: 23
Conservative: 7
Mismatches: 13
Indels: 0
Gaps: 0

US-10-045-180A-5 (1-44) x AA722853 (1-541)

Qy   2 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 21
          ||||||||| :::||||:||| ||::||| :::|||||||
Db   431 CCACTCCGGCCAAGAGGTGATGAGGCTCCAGGCCAGGCGTGGGCCAGAAGACCAG 372

Qy   22 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 41
          |||:|||||||||||||||||| :::|||||||||||||
Db   371 GACATATCTATTTCCTTTGCATGGGATAAAAGCTCTGCTTCTTCAGGTTCAACA 312

Qy   42 LysGlyLeu 44
          :::||:|||
Db   311 AGGGGCATG 303

RESULT 5

```
AI250799/c
LOCUS           AI250799        542 bp    mRNA    linear    EST 28-JAN-1999
DEFINITION      qi36g07.x1 Soares NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1858620
                3', similar to gb:X65977 NEUTROPHIL DEFENSIN 4 PRECURSOR (HUMAN);,
                mRNA sequence.
ACCESSION       AI250799
VERSION         AI250799.1  GI:3847328
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 542)
  AUTHORS       NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE         National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
  JOURNAL       Unpublished
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                This clone is available royalty-free through LLNL ; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                Seq primer: -40UP from Gibco
                High quality sequence stop: 463.
                Insert Length: 633   Std Error: 0.00
FEATURES             Location/Qualifiers
     source          1..542
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /clone="IMAGE:1858620"
                     /db_xref="taxon:9606"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /clone_lib="Soares NhHMPu_S1"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site 1: Not I;
                     Site 2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     34048B-345679, and 484488-489479."
                     /lab_host="DH10B"
BASE COUNT      151 a     134 c     136 g     119 t     2 others
ORIGIN
```

US-10-045-180A-5 (1-44) x AI250799 (1-542)

Alignment Scores:
Pred. No.:
Score: 114.00   Length: 542
Percent Similarity: 69.77%   Matches: 23
Best Local Similarity: 53.49%   Conservative: 7
Query Match: 9   Mismatches: 13
DB: Indels: 0
                                    Gaps: 0

```
Qy      2 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 21
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    431 CCACTCCAGGCAAGAGAGTGATGAGGCTCCAGGCCAGGAAGACCAG 372

Qy     22 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 41
          ||||||||:::  ||| |||||:||||||||||||||||||||||||||||||||||||
Db    371 GACATATCTATTCCTTGCATGGGATAAAAGCTTCTGCTCTTCAGGTTTCAGGCTCAACA 312

Qy     42 LysGlyLeu 44
          :::|||||
Db    311 AGGGGCATG 303
```

RESULT 6

```
BU616655/c
LOCUS           BU616655        634 bp    mRNA    linear    EST 23-SEP-2002
DEFINITION      UI-H-DF0-ben-n-14-0-UI.s1 NCI_CGAP_DF0 Homo sapiens cDNA clone
                UI-H-DF0-ben-n-14-0-UI 3', mRNA sequence.
ACCESSION       BU616655
VERSION         BU616655.1  GI:23282870
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 634)
  AUTHORS       NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE         National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
  JOURNAL       Unpublished
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: Dr. Jose Mercuende
                cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
                cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
                DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
                Clone Distribution: Clone distribution information can be obtained
                from Dr. M. Bento Soares, bento-soares@uiowa.edu
                Seq primer: M13 FORWARD
FEATURES             Location/Qualifiers
     source          1..634
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /clone="UI-H-DF0-ben-n-14-0-UI"
                     /db_xref="taxon:9606"
                     /tissue_type="Subchondral Bone"
                     /dev_stage="Adult"
                     /clone_lib="NCI_CGAP_DF0 (Life Technologies)"
                     /note="Organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site 1: EcoR I; Site 2: Not I;
                     NCI_CGAP_DF0 is a cDNA library containing the following
                     tissue(s): Subchondral Bone. The library was constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is GTTAAGCGTC.
                     TAG_LIB=UI-H-DF0
                     TAG_TISSU=subchondral bone
                     TAG_SEQ=GTTAAGCGTC"
                     POLYA=yes.
BASE COUNT      170 a     151 c     158 g     155 t
ORIGIN
```

US-10-045-180A-5 (1-44) x BU616655 (1-634)

Alignment Scores:
Pred. No.:
Score: 114.00   Length: 634
Percent Similarity: 69.77%   Matches: 23
Best Local Similarity: 53.49%   Conservative: 7
Query Match: 13   Mismatches: 13
DB: Indels: 0
                                    Gaps: 0

```
Qy      2 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 21
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    449 CCACTCCAGGCAAGAGAGTGATGAGGCTCCAGGCCAGGAAGACCAG 390

Qy     22 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 41
          ||||||||:::  ||| |||||:||||||||||||||||||||||||||||||||||||
Db    389 GACATATCTATTCCTTGCATGGGATAAAAGCTTCTGCTCTTCAGGTTTCAGGCTCAACA 330

Qy     42 LysGlyLeu 44
```

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: December 17, 2003, 14:50:33 ; Search time 165 Seconds
(without alignments)
719.850 Million cell updates/sec

Title: US-10-045-180A-5
Perfect score: 233
Sequence: 233
1 EPLQARAHEMPAQKQPPADD..........IYFSGDDSCSLQVPGSTKGL 44

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop   6.0 , Delext 7.0

Searched: 252756 seqs, 134719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
```
-MODEL=frame+_p2n.model
-DEV=xlp
-Q=/cgn2_1/USPTO_epool_p/US10045180/runat_17122003_145023_12520/app_query.fasta_1.860
-DB=N_Geneseq_19Jun03  -QFMT=fastap  -SUFFIX=rng  -MINMATCH=0.1  -LOOPCL=0
-LOOPEXT=0  -UNITS=bits  -START=1  -END=-1  -MATRIX=blosum62  -TRANS=human40.cdi
-LIST=45  -DOCALIGN=200  -THR_MAX=100  -THR_MIN=0  -ALIGN=15
-MODE=LOCAL  -OUTFMT=pto  -NORM=ext  -HEAPSIZE=500  -MINLEN=0  -MAXLEN=200000000
-USER=US10045180 @CGN 1_1_835 @runat_17122003_145023_12520  -ICPU=3
-NO_NMAP  -LARGEQUERY_NEG_SCORES=0  -WAIT  -DSPBLOCK=100  -LONGLOG
-DEV_TIMEOUT=120  -WARN_TIMEOUT=30  -THREADS=1  -XGAPOP=10  -XGAPEXT=0.5  -FGAPOP=6
-FGAPEXT=7  -YGAPOP=10  -YGAPEXT=0.5  -DELOP=6  -DELEXT=7
```

Database :
```
N_Geneseq_19Jun03:*
 1: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1980.DAT:*
 2: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1981.DAT:*
 3: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1982.DAT:*
 4: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1983.DAT:*
 5: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1984.DAT:*
 6: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1985.DAT:*
 7: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1986.DAT:*
 8: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1987.DAT:*
 9: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	233	100.0	453	20	AAX26697	cDNA sequence enco
2	233	100.0	453	20	AAX26698	cDNA sequence enco
3	209	89.7	4415	20	AAX26696	Nucleotide sequence
4	114	48.9	860	25	ABX63473	Human cDNA #473 di
5	104	44.6	496	25	AAD51767	Human retrocyclin
6	101	43.3	294	22	AAH57220	Human intestine sp
7	101	43.3	452	14	AAQ53217	Sequence encoding
8	101	43.3	452	25	ACA56600	Human signaling p
9	90	38.6	398	24	ABN59997	Novel human coding
10	86	36.9	243	22	ABQ60783	Human defensin 5 D
11	83	35.6	495	22	AAC66925	Rhesus macaque the
12	83	35.6	495	25	AAD51769	Macaca mulatta RTD
13	82	35.2	500	22	AAC66924	Rhesus macaque the
14	82	35.2	500	25	AAD51768	Macaca mulatta RTD
15	80	34.3	424	14	AAQ53216	Sequence encoding
16	80	34.3	449	24	ABQ60784	Human defensin 5 D
17	80	34.3	514	22	AAH57427	Human intestine ce
18	80	34.3	586	24	ABQ60212	Human colon cancer
c 19	80	34.3	2548	22	AAC66927	Rhesus macaque the
20	78.5	33.7	299	16	AAF26106	Human defensin 3 P
21	78.5	33.7	464	21	AAF20911	Human adenosine re
22	78.5	33.7	464	21	AAA34789	Human defensin 3 P
23	78.5	33.7	478	21	AAC04468	Human secreted pro
24	78.5	33.7	498	12	AAQ01908	Chronic myelogenou
25	78.5	33.7	498	24	ABQ80601	Human defensin 1 c
26	78.5	33.7	514	24	ABN59722	Novel human coding
27	78.5	33.7	543	25	ABZ36767	Human GENSET codin
28	78.5	33.7	644	21	AAF21446	Human defensin 3 p
29	78.5	33.7	652	24	ABZ11202	Human polynucleoti
30	78.5	33.7	664	24	ABZ11183	Human polynucleoti
31	76	33.3	209273	21	AAF21437	Human factor-relat
32	76	32.6	2523	22	AAC66926	Rhesus macaque the
33	76	32.6	2880	14	AAQ53218	Genomic sequence e
34	74	31.8	445	17	AAT30734	Rat cryptdin 2 CDN
35	74	31.8	2408	17	AAT30737	Rat cryptdin 2 gen
36	73.5	31.5	411	24	ABQ60009	Human colon cancer
37	73	31.3	473	25	ABZ36740	Human colon cancer
38	72.5	31.1	652	24	ABZ12202	Human polynucleoti
39	72	30.9	2457	17	AAT30736	Rat cryptdin 1 gen
40	72	30.9	445	17	AAT30738	Rat cryptdin 3 gen
41	71	30.5	2551	17	AAT30735	Rat cryptdin 3 CDN
42	71	30.5	445	17	ABK30738	Rat cryptdin 3 CDN
43	70	30.0	445	24	AAT47375	Human LP321 DNA.
c 44	69	29.6	480	24	ABK39009	cDNA encoding lung
c 45	69	29.6	412	25	ACA11338	Human lung adenoca

RESULT 1
AAX26697

```
ID   AAX26697 standard; cDNA; 453 BP.
XX
AC   AAX26697;
XX
DT   18-JUN-1999  (first entry)
XX
DE   cDNA sequence encoding human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis; ss.
XX
OS   Homo sapiens.
```

ALIGNMENTS

Left column

```
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;   97FR-0010823.
XX
PR   29-AUG-1997;   97FR-0010823.
XX
PI   Bougueleret L, Chumakov I;
XX
PA   (GEST ) GENSET SA.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
     anticancer agent, pesticide, etc.
XX
DR   WPI; 1999-183266/16.
DR   P-PSDB; AAY01604.
XX
PS   Disclosure; Fig 3; 56pp; French.
XX
CC   The present sequence encodes human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 453 BP; 107 A; 121 C; 97 G; 128 T; 0 other;
```

Alignment Scores:

Pred. No.:	
Score:	2.62e-25
Percent Similarity:	233.00
Best Local Similarity:	100.00%
Query Match:	100.00%

Length:	453
Matches:	44
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-045-180A-5 (1-44) x AAX26697 (1-453)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnIlyLysGlnProAlaAlaAspAsp 20
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 109 GAGCCGCTCCAGGCAAGGCAAGAGCTCATGAGAATGCCAGCCCAGAAGCAGCCTTCCAGCAGATGAC 168

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
         ||||||||||||||||||||||||||||||||||||||||
Db 169 CAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGCTCA 228

Qy  41 ThrLysGlyLeu 44
         ||||||||||||
Db 229 ACAAAGGGCTTG 240
```

```
RESULT 2
AAX26698
AC   AAX26698;
ID   AAX26698 standard; cDNA; 453 BP.
XX
DT   18-JUN-1999 (first entry)
XX
DE   cDNA sequence encoding human defensin (Def-X) protein.
```

Right column

```
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;   97FR-0010823.
XX
PR   29-AUG-1997;   97FR-0010823.
XX
PI   Bougueleret L, Chumakov I;
XX
PA   (GEST ) GENSET SA.
XX
DR   WPI; 1999-183266/16.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
     anticancer agent, pesticide, etc.
XX
PS   Claim 5; Fig 4; 56pp; French.
XX
CC   The present sequence encodes human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 453 BP; 107 A; 121 C; 97 G; 128 T; 0 other;
```

Alignment Scores:

Pred. No.:	
Score:	2.62e-25
Percent Similarity:	233.00
Best Local Similarity:	100.00%
Query Match:	100.00%

Length:	453
Matches:	44
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-045-180A-5 (1-44) x AAX26698 (1-453)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnIlyLysGlnProAlaAlaAspAsp 20
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 109 GAGCCGCTCCAGGCAAGGCAAGAGCTCATGAGAATGCCAGCCCAGAAGCAGCCTTCCAGCAGATGAC 168

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
         ||||||||||||||||||||||||||||||||||||||||
Db 169 CAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGCTCA 228

Qy  41 ThrLysGlyLeu 44
         ||||||||||||
Db 229 ACAAAGGGCTTG 240
```

```
RESULT 3
AAX26696
```

ID   AAX26696 standard; DNA; 4415 BP.
XX
AC   AAX26696;
XX
DT   18-JUN-1999 (first entry)
XX
DE   Nucleotide sequence of human defensin (Def-X).
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;   97FR-0010823.
XX
PR   29-AUG-1997;   97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Claim 5; Fig 2; 56pp; French.
XX
CC   The present sequence encodes human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 4415 BP; 1128 A; 1109 C; 912 G; 1252 T; 14 other;

Alignment Scores:

Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:

Alignment Scores:
  Pred. No.:   1.74e-20
  Score: 209.00   Length: 4415
  Percent Similarity: 100.00%   Matches: 39
  Best Local Similarity: 100.00%   Conservative: 0
  Query Match: 89.70%   Mismatches: 0
  DB: 20   Indels: 0
                                      Gaps: 0

US-10-045-180A-5 (1-44) x AAX26696 (1-4415)

Qy   1 GluProLeuGlnIleAlaArgAlaHisGlyMetProAlaGlnInlysGlnProProAlaAspAsp 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3463 GAGCCGCTCCAGGCAAGAGCTCATCAGATCCACGCCCAGAAGCAGCCTCCAGCAGATGAC 3522

Qy   21 GlnAspValAlaIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly 39
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3523 CAGGATGTGGTCATTTACTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCAGGT 3579

RESULT 4
ABX63473
XX
AC   ABX63473;
XX
ID   ABX63473 standard; cDNA; 860 BP.
XX
DT   26-FEB-2003 (first entry)
XX
DE   Human cDNA #473 differentially expressed in activated vascular tissue.
XX
KW   Human; gene; ss; vascular tissue; cytostatic; atherosclerosis;
KW   cardiant; hypotensive; antidiabetic; gynaecological; vasotropic;
KW   cerebroprotective; gene therapy; vascular disease; cancer; coronary;
KW   artery disease; hypertension; diabetes; pre-eclampsia; restenosis;
KW   ischaemia-reperfusion injury; stroke;
XX
OS   Homo sapiens.
XX
PN   US2002137081-A1.
XX
PD   26-SEP-2002.
XX
PF   08-JAN-2002; 2002US-0044090.
XX
PR   28-JUL-2000; 2000US-222469P.
PR   08-JAN-2001; 2001US-260483P.
XX
PA   (BAND/) BANDMAN O.
XX
PI   Bandman O;
XX
DR   WPI; 2003-110597/10.
XX
PT   Combination for diagnosing, staging, treating, or monitoring the
PT   progression of treatment of a vascular disease, e.g. atherosclerosis,
PT   comprises several cDNAs that are differentially expressed in activated
PT   vascular tissue -
XX
PS   Claim 1; Page -; 18pp; English.
XX
CC   This invention relates to a combination comprising several cDNAs that
CC   are differentially expressed in activated vascular tissue. The invention
CC   also discloses a high throughput method for detecting differentially
CC   expressed cDNAs in a sample. The cDNAs of the invention may have
CC   antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC   gynaecological; vasotropic and cerebroprotective activities and may be
CC   used in gene therapy. The cDNAs of the invention may be used in a
CC   high-throughput methods for detecting differential expression of one or
CC   more cDNAs in a sample, or screening several molecules or compounds to
CC   identify a molecule or compound that specifically binds a cDNA of the
CC   invention. A protein encoded by the cDNA may be used to screen several
CC   molecules or compounds to identify a ligand that specifically binds to
CC   the protein, or to produce or purify a protein in a sample or purify a natural or
CC   recombinant protein from a sample. The nucleotides may be useful for
CC   be used to detect a protein in a sample or purify a natural or
CC   diagnosing, staging, treating, or monitoring the progression of
CC   treatment of a vascular disease, e.g. atherosclerosis, cancer, coronary
CC   artery disease, hypertension, diabetes, pre-eclampsia, ischaemia-
CC   reperfusion injury, restenosis, or stroke. The cDNAs can also be used
CC   for large-scale genetic or gene expression analysis of several new
CC   nucleic acid molecules. Antibodies to the proteins encoded by the
CC   cDNAs are useful for diagnosing pre-pathologic disorders, and chronic
CC   or acute diseases associated with abnormalities in the expression,
CC   amount or distribution of the invention that is differentially expressed in
CC   represents a cDNA of the invention that is differentially expressed in
CC   activated vascular tissue.
CC   Note: The sequence data for this patent did not form part of the
CC   specification, but was obtained in electronic format directly from USPTO
CC   at http.seqdata.uspto.gov/sequence.html?DocID=20020137081.
XX
SQ   Sequence 860 BP; 216 A; 211 C; 192 G; 241 T; 0 other;

Left column

Alignment Scores:
```
Pred. No.:              2.78e-07
Score:                  114.00      Length:            860
Percent Similarity:     69.77%      Matches:           23
Best Local Similarity:  53.49%      Conservative:      7
Query Match:            48.93%      Mismatches:        13
                        25          Indels:            0
DB:                                 Gaps:              0
```

US-10-045-180A-5 (1-44) x ABX63473 (1-860)

```
Qy    2  ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln  21
         ||||||||||||||||||||||||||:||:||||          :::||||||
Db  136  CCACTCCAGGCAAGAGGTGATGAGGCTGCTCAGGCCACCGTGGGCCCAGAAGACCAG 195

Qy   22  AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr  41
         |||::|          |||||:::||||||||||||||::||||
Db  196  GACATAATCTATTTCCTTTGCATGGGATAAAAAGCTTCGTGTTCTTCAGGCTTCAGGCTCAACA 255

Qy   42  LysGlyLeu  44
         :::||:||:
Db  256  AGGGGCATG  264
```

RESULT 5
```
AAD51767
ID   AAD51767 standard; DNA; 496 BP.
XX
AC   AAD51767;
XX
DT   16-APR-2003  (first entry)
XX
DE   Human retrocyclin DNA.
XX
KW   Retrocyclin; infection; sexually transmitted disease; gene therapy; HIV;
KW   human immunodeficiency virus; bacterial vaginosis; ophthalmic infection;
KW   antibiotic modelling; antimicrobial; human; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
CDS  63..495
FT                   /partial
FT                   /product= "Human retrocyclin precursor protein"
FT                   /tag= a
FT   sig_peptide     63..126
FT                   /*tag= b
FT   mat_peptide     127..495
FT                   /*tag= c
FT   CDS             127..306
FT                   /product= "Mature human retrocyclin protein"
FT                   /*tag= d
FT                   /note= "Xaa corresponds to in-frame stop codon;
FT                   No start and stop codon;
FT                   /product= "Human mature retrocyclin protein fragment"
XX
PN   WO200285401-A1.
XX
PD   31-OCT-2002.
XX
PF   18-APR-2002; 2002WO-US12353.
XX
PR   18-APR-2001; 2001US-28485SP.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Lehrer RI, Waring AJ, Cole AM, Hong TB;
XX
DR   WPI; 2003-103387/09.
DR   P-PSDB; AAB33811, AAB33860.
XX
```

Right column

```
PT   New isolated retrocyclin peptide, useful for preventing retroviral
PT   infections in cells susceptible to bacterial or viral infections or
PT   treating patients having the infections, such as HIV, sexually
PT   transmitted diseases, vaginosis
XX
PS   Claim 14; Page 50-51; 72pp; English.
XX
CC   The invention relates to novel retrocyclin peptides. Peptides and
CC   methods of the invention are useful for preventing retroviral infections
CC   in cells susceptible to the invention or bacterial or viral infections, or treating
CC   patients having infections such as HIV (human immunodeficiency virus),
CC   sexually transmitted diseases, bacterial vaginosis or ophthalmic
CC   infections. The retrocyclin-mediated killing is useful for modelling
CC   and screening novel antibiotics. The invention is also useful in gene
CC   therapy. The present sequence is human retrocyclin DNA.
XX
SQ   Sequence 496 BP; 106 A; 138 C; 117 G; 135 T; 0 other;
```

US-10-045-180A-5 (1-496)

Alignment Scores:
```
Pred. No.:              4.22e-06
Score:                  104.00      Length:            496
Percent Similarity:     68.18%      Matches:           22
Best Local Similarity:  50.00%      Conservative:      8
Query Match:            44.64%      Mismatches:        14
                        25          Indels:            0
DB:                                 Gaps:              0
```

```
Qy    1  GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp  20
         |||||||||||||||||||||||||||||||||||||||:||:|||||
Db  136  GAGCCACTTCAGGCAAGAGCTGATGAGGCTGCAGCCCAGGAGCAGCCTGGAGCAGATGAT 195

Qy   21  AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer  40
         |||::|          |||||:::|||||       :::||||       :::|||
Db  196  CAGGAAAATGGCTCATGCCTTTACATGGCATGAAAAGTGCCGGCCGCTTCCGCTTCAGACTCA 255

Qy   41  ThrLysGlyLeu  44
         :::||||||
Db  256  GCGAGGAGCTTG 267
```

RESULT 6
```
AAH57220
ID   AAH57220 standard; cDNA; 294 BP.
XX
AC   AAH57220;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human intestine specific cDNA sequence SEQ ID NO:60.
XX
KW   Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
KW   lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW   metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW   neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX
OS   Homo sapiens.
XX
PN   WO200132927-A2.
XX
PD   10-MAY-2001.
XX
PF   02-NOV-2000; 2000WO-US30396.
XX
PR   04-NOV-1999; 99US-0163508.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Sornasse T, Seilhamer JJ, Watson GA;
XX
DR   WPI; 2001-291057/30.
XX
PT   New cell and tissue specific polynucleotides useful for diagnosis,
PT   prognosis or monitoring of treatments for disorders where the gene is
```

Column 1

```
PT   associated with a cancer, immunopathology or neuropathology -
XX
PS   Claim 1; Page 95; 327pp; English.
XX
CC   AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
CC   sequences (I). (I) can have cytotoxic, immunomodulatory and
CC   neuroprotective activities, and can be used in gene therapy. (I) and
CC   proteins (II) encoded by then are used in high throughput screening
CC   assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC   mimetics, peptides, proteins, agonists, antagonists, antibodies or
CC   their fragments, immunoglobulins, inhibitors, drug compounds and
CC   pharmaceutical agents. Expression of (I) in a sample indicates the
CC   differentiation of embryonic stem cells into a tissue selected from
CC   brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC   tissues. (I) and (II) are used to produce an expression profile that
CC   defines a metabolic or developmental process, treatment, condition,
CC   disease or disorder. The gene profile can be used for diagnosis,
CC   prognosis or monitoring of treatments and for investigating a
CC   predisposition to a disorder where the gene is associated with a
CC   cancer, immunopathology or neuropathology.
XX
SQ   Sequence 294 BP; 72 A; 80 C; 73 G; 64 T; 5 other;
```

US-10-045-180A-5 (1-44) x AAH57220 (1-294)

```
Alignment Scores:
Pred. No.:
Score:                   6.04e-06     Length:          294
Percent Similarity:      101.00       Matches:          21
Best Local Similarity:   71.43%       Conservative:      9
Query Match:             50.00%       Mismatches:       12
                         43.35%       Indels:            0
                         22           Gaps:              0

Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
        :::||||||||||||||||::||||||||||||||:::||||||||:::
Db   93 GATCCACTGCAGGCAAAAAGCTTATGAGGCTGATGCCCAGGAGCAGCCGGCCAAATGAC 152

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
        |||||  :::  |||:::   |||||||||||:::
Db  153 CAGGACTTTGCCGTCTCCTTTGCAGAGGATGCAAGCTCAAGTCTTAGAGCTTTGGGCTCA 212

Qy   41 ThrLys 42
        |||:::
Db  213 ACAAGG 218
```

Column 2

```
XX
CC   A probe (D5' oligo) having the sequence given in AAQ53219 (claimed)
CC   was used to screen a population of clones to identify novel defensin
CC   defensin peptides. Pref. the libraries were human genomic and cDNA
CC   libraries. Hybridisation and partial sequence analysis of the
CC   identified clones contained previously characterised myeloid derived
CC   defensin sequences as well as new defensin related sequences. Two
CC   clones expressing new defensin related sequences were extensively
CC   characterised and found to contain genes selectively expressed in
CC   Paneth cells of the small intestin. These Paneth cell-derived
CC   defensins are designated human defensin 5 and human defensin 6 and
CC   are referred to as gastrointestinal defensin peptides. Comparison
CC   of the deduced AA sequences of the defensin 5 cDNA with the
CC   previously reported preprodefensin shows significant similarity.
CC   The deduced AA sequence of defensin 6 cDNA has features similar to
CC   defensin 5 and the previously reported preprodefensins.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 452 BP; 123 A; 118 C; 98 G; 113 T; 0 other;
```

US-10-045-180A-5 (1-44) x AAQ53217 (1-452)

```
Alignment Scores:
Pred. No.:
Score:                   1.04e-05     Length:          452
Percent Similarity:      101.00       Matches:          21
Best Local Similarity:   71.43%       Conservative:      9
Query Match:             50.00%       Mismatches:       12
                         43.35%       Indels:            0
                         14           Gaps:              0

Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
        :::||||||||||||||||::||||||||||||||:::||||||||:::
Db   97 GATCCACTGCAGGCAAAAAGCTTATGAGGCTGATGCCCAGGAGCAGCCGGCCAAATGAC 156

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
        |||||  :::  |||:::   |||||||||||:::
Db  157 CAGGACTTTGCCGTCTCCTTTGCAGAGGATGCAAGCTCAAGTCTTAGAGCTTTGGGCTCA 216

Qy   41 ThrLys 42
        |||:::
Db  217 ACAAGG 222
```

Left column

```
XX
PD   31-DEC-2002.
XX
PF   30-JAN-1998;   98US-0016434.
XX
PR   30-JAN-1998;   98US-0016434.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Au-Young J, Seilhamer JJ;
XX
DR   WPI; 2003-352189/33.
XX
PT   Combination of polynucleotide probes, useful as array elements in a
PT   microarray for monitoring the expression of a number of target
PT   polynucleotides -
XX
PS   Claim 1; SEQ ID NO 1198; 65pp; English.
XX
CC   The invention relates to a combination which, comprises a number of
CC   polynucleotide probes comprising a sequence selected from one of the 1490
CC   sequences mentioned in the specification. The combination is useful as an
CC   array element in a microarray for monitoring the expression of a number
CC   of target polynucleotides. The microarray is particularly useful in the
CC   diagnosis and treatment of cancer and immunopathology and neuropathology.
CC   The microarray is useful in diagnostics and treatment regimens, drug
CC   discovery and development, toxicological and carcinogenicity studies,
CC   forensics and pharmacogenomics. The microarray is also useful for
CC   monitoring progression of diseases and for developing sophisticated
CC   profiles for the effects of currently available therapeutic drugs. The
CC   combination is also useful for purifying a subpopulation of mRNAs, cDNAs
CC   and genomic fragments and in research and diagnostic applications. The
CC   array can detect changes in expression in a large number of genes coding
CC   for different signaling pathway populations which can be used to diagnose
CC   various diseases including cancer e.g. adenocarcinoma and leukaemia,
CC   immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
CC   and Parkinson's disease. The present sequence represents a polynucleotide
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=0650093B1.
XX
SQ   Sequence 452 BP; 124 A; 118 C; 98 G; 112 T; 0 other;
XX

US-10-045-180A-5 (1-44) x ACA56600 (1-452)

Alignment Scores:
   Pred. No.:
   Score:                  1.0e-05
   Percent Similarity:     101.00          Length:        452
   Best Local Similarity:  71.43%          Matches:       21
   Query Match:            50.00%          Conservative:  9
                           43.35%          Mismatches:    12
                           25              Indels:        0
                                           Gaps:          0

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlyGlnLysGlnProProAlaAspAsp 20
       ::||||||||||||||||||||||| :||:||||||||||||||||::||||||||||:||||
Db  97 GATCCACTGCAGGCAAAAGCTTATGAGGGCTGAGTGCCCAGAGACCAGCGTGGGGCAAATGAC 156

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
       |||||||||||||:::||||::|||||||||||||||||||||||||||:::|| |||||
Db 157 CAGGACTTTGCCGTCTCCTTTGCAGGAGATGCAAGCTCAAGTCTTAGAGCTTTGGGCTCA 216

Qy  41 ThrLys 42
       ||:|||
Db 217 ACAAGG 222
```

```
RESULT 9
ABNS9997
ID   ABNS9997 standard; cDNA, 398 BP.
XX
AC   ABNS9997;
XX
DT   28-JUN-2002 (first entry)
XX
```

Right column

```
XX
DE   Novel human coding sequence SEQ ID NO: 408.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   11-SEP-2000; 2000US-0659671.
XX
PR   10-SEP-2001; 2001WO-US26015.
XX
PA   (HYSR-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drnanac RT;
XX
DR   WPI; 2002-292408/33.
DR   P-PSDB; ABB97584.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Claim 1; SEQ ID NO 408; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC   e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a coding sequence of the
CC   invention.
XX
SQ   Sequence 398 BP; 102 A; 108 C; 93 G; 95 T; 0 other;
XX

US-10-045-180A-5 (1-44) x ABNS9997 (1-398)

Alignment Scores:
   Pred. No.:
   Score:                  0.000382
   Percent Similarity:     90.00           Length:        398
   Best Local Similarity:  69.23%          Matches:       19
   Query Match:            48.72%          Conservative:  8
                           38.63%          Mismatches:    12
                           24              Indels:        0
                                           Gaps:          0

Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlyGlnLysGlnProProAlaAspAsp 20
        :::||||||||||||||||||||| :||:||||||||||||||::||||||||||||:||||
Db  108 GATCCACTGCAGGCAAAAGCTTATGAGGGCTGAGTGCCCAGAGACCAGCGTGGGGCAAATGAC 167

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly 39
        |||||||||||||:::||||::||||||||||||||||||||||||||:::||||||
Db  168 CAGGACTTTGCCGTCTCCTTTGCAGGAGATGCAAGCTCAAGTCTTAGAGCTTTGGGT 224
```

```
RESULT 10
AAC66930
ID   AAC66930 standard; cDNA, 243 BP.
XX
AC   AAC66930;
XX
DT   27-MAR-2001 (first entry)
XX
DE   Human theta defensin coding sequence SEQ ID NO: 28.
XX
```

KW   Theta defensin; antimicrobial; cyclic; bacterium; fungus; protozoan;
KW   virus; helminth; disinfectant; food preservative; analogue; ss.
XX
OS   Homo sapiens.
XX
PN   WO200068265-A1.
XX
PD   16-NOV-2000.
XX
PF   10-MAY-2000; 2000WO-US12842.
XX
PR   10-MAY-1999; 99US-0309487.
XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Selsted ME, Tang Y, Yuan J, Ouellette AJ;
XX
DR   WPI; 2001-031853/04.
XX
PT   Novel theta defensin peptide with antimicrobial activity against
CC   bacteria, yeast, fungi, protozoa and viruses -
XX
PS   Example 5; Fig 15; 110pp; English.
XX
CC   The present invention provides theta defensin peptides and analogues
CC   which have antimicrobial activity. They can be used in the treatment of
CC   bacterial, viral, fungal, protozoan and helminthic infections, in
CC   disinfectants and as food preservatives.
XX
SQ   Sequence 243 BP; 52 A; 75 C; 65 G; 51 T; 0 other;

US-10-045-180A-5 (1-44) x AAC66930 (1-243)

Alignment Scores:
Pred. No.:
Score: 0.0008
                          86.00
Percent Similarity: 61.36%   Length: 243
Best Local Similarity: 45.45%   Matches: 20
Query Match: 36.91%   Conservative: 7
DB: 22   Mismatches: 17
                                        Indels: 0
                                        Gaps: 0

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
          ||||||||||||||||||||||||||||||||||||::||||||||||||||||||||
Db   67 GAGGCACGTCAGGCAAGAGCTGATGAAGAGCTGCCCCCAGCAGCAGCCTGGAGCAGATGAT 126

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
          |||::| :|||||| ||||||
Db 127 CAGGGAATGGCTCATTCCTTTACTGGCTGAAAAACGCCGGCTCTTCCACTTTCAGAGTCA 186

Qy   41 ThrLysGlyLeu 44
          ||||||||||||
Db 187 GCCGAAAGGCTTG 198

RESULT 11
ID   AAC66925 standard; cDNA; 495 BP.
XX
AC   AAC66925;
XX
DT   27-MAR-2001 (first entry)
XX
DE   Rhesus macaque theta defensin RTD-1b coding sequence.
XX
KW   Theta defensin; antimicrobial; cyclic; bacterium; fungus; protozoan;
KW   virus; helminth; disinfectant; food preservative; analogue; ss.
XX
OS   Rhesus macaque.
XX
PN   WO200068265-A1.
XX
PD   16-NOV-2000.
XX
PF   10-MAY-2000; 2000WO-US12842.
XX

XX
PR   10-MAY-1999; 99US-0309487.
XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Selsted ME, Tang Y, Yuan J, Ouellette AJ;
XX
DR   WPI; 2001-031853/04.
DR   P-PSDB; AAB35041.
XX
PT   Novel theta defensin peptide with antimicrobial activity against
CC   bacteria, yeast, fungi, protozoa and viruses -
XX
PS   Example 5; Fig 11; 110pp; English.
XX
CC   The present invention provides theta defensin peptides and analogues
CC   which have antimicrobial activity. They can be used in the treatment of
CC   bacterial, viral, fungal, protozoan and helminthic infections, in
CC   disinfectants and as food preservatives.
XX
SQ   Sequence 495 BP; 103 A; 145 C; 122 G; 125 T; 0 other;

US-10-045-180A-5 (1-44) x AAC66925 (1-495)

Alignment Scores:
Pred. No.:
Score: 0.0553
                          83.00
Percent Similarity: 61.36%   Length: 495
Best Local Similarity: 43.18%   Matches: 19
Query Match: 35.62%   Conservative: 8
DB: 22   Mismatches: 17
                                        Indels: 0
                                        Gaps: 0

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
          ||||||||||||||||||||||||||||||||||||::||||||||||||||||||||
Db 150 GAGGCACGTCAGGCAAGAGCTGATGAAGAGCTGCCCCCAGCAGCAGCCTGGAGCAGATGAT 209

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
          |||::| :|||||| ||||||
Db 210 CAGGGAATGGCTCATTCCTTTACTGGCTGAAAAACGCCGGCTCTTCCGGCGTTTCAGAGTCA 269

Qy   41 ThrLysGlyLeu 44
          ::|||||||||
Db 270 GCGAGAGGCTTG 281

RESULT 12
ID   AAD51769 standard; DNA; 495 BP.
XX
AC   AAD51769;
XX
DT   16-APR-2003 (first entry)
XX
DE   Macaca mulatta RTD2 DNA.
XX
KW   Retrocyclin; infection; sexually transmitted disease; gene therapy; HIV;
KW   human immunodeficiency virus; bacterial vaginosis; ophthalmic infection;
KW   antibiotic modelling; antimicrobial; rhesus monkey; theta defensin 1b;
KW   RTD2; gene; ds.
OS   Macaca mulatta.
XX
FH   Key   Location/Qualifiers
FT   CDS   90..320
FT   /*tag= a
FT   /product= "RTD2 protein"
FT   sig_peptide   90..149
FT   /*tag= b
FT   mat_peptide   150..317
FT   /*tag= c
FT   /product= "Mature RTD2 protein"
XX
PN   WO200285401-A1.
XX

Left column

```
PD   31-OCT-2002.
XX
PF   18-APR-2002; 2002WO-US12353.
XX
PR   18-APR-2001; 2001US-284855P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Lehrer RI, Waring AJ, Cole AM, Hong TB;
XX
DR   WPI; 2003-103387/09.
DR   P-PSDB; AAE33813.
XX
PS   Disclosure; Page 53; 72pp; English.
XX
PT   New isolated retrocyclin peptide, useful for preventing retroviral
PT   infections in cells susceptible to bacterial or viral infections or
PT   treating patients having the infections, such as HIV, sexually
PT   transmitted diseases, vaginosis -
XX
CC   The invention relates to novel retrocyclin peptides. Peptides and
CC   methods of the invention are useful for preventing retroviral infections
CC   in cells susceptible to bacterial or viral infections, or treating
CC   patients having infections such as HIV (human immunodeficiency virus),
CC   sexually transmitted diseases, bacterial vaginosis or ophthalmic
CC   infections. The retrocyclin-mediated killing is useful for modelling
CC   and screening novel antibiotics. The invention is also useful in gene
CC   therapy. The present sequence is rhesus monkey theta defensin 1b
CC   precursor, RTD2 DNA. This sequence is used in the exemplification of
CC   the invention.
XX
SQ   Sequence 495 BP; 103 A; 145 C; 122 G; 125 T; 0 other;
```

US-10-045-180A-5 (1-44) x AAD51769 (1-495)

Alignment Scores:

Pred. No.:	0.0553	Length: 495
Score:	83.00	Matches: 19
Percent Similarity:	61.36%	Conservative: 8
Best Local Similarity:	43.18%	Mismatches: 17
Query Match:	25	Indels: 0
		Gaps: 0

```
Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlyHisGlnProProAlaAspAsp 20
         |||||||||:|||||||||||||:::||||||:|||||||||||||||||||
Db  150 GAGGCACGTCAGGCAAGAGCTGATGAAGCTCCGCCCAGCAGCCTGGAGCAGATGAT 209

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
         ||||::  :::|||::  :::|||
Db  210 CAGGGAAATGGCTCATTCCTTTACACGGCCTGAAAACGCCCTCTTCCGCTTTCAGAGTCA 269

Qy   41 ThrLysGlyLeu 44
         ::|||||||||
Db  270 GCGGAGGCTTG 281
```

```
RESULT 13
AAC66924
ID   AAC66924 standard; cDNA; 500 BP.
XX
AC   AAC66924;
XX
DT   27-MAR-2001 (first entry)
XX
DE   Rhesus macaque theta defensin RTD-1a coding sequence.
XX
KW   Theta defensin; antimicrobial; cyclic; bacterium; fungus; protozoan;
KW   virus; helminth; disinfectant; food preservative; analogue; ss.
XX
OS   Rhesus macaque.
XX
PN   WO200068265-A1.
XX
PD   16-NOV-2000.
```

Right column

```
XX
PF   10-MAY-2000; 2000WO-US12842.
XX
PR   10-MAY-1999; 99US-0309487.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Selsted ME, Tang Y, Yuan J, Ouellette AJ;
XX
DR   WPI; 2001-031853/04.
DR   P-PSDB; AAB35040.
XX
PS   Example 5; Fig 11; 110pp; English.
XX
PT   Novel theta defensin peptide with antimicrobial activity against
PT   bacteria, yeast, fungi, protozoa and viruses -
XX
CC   The present invention provides theta defensin peptides and analogues
CC   which have antimicrobial activity. They can be used in the treatment of
CC   bacterial, viral, fungal, protozoan and helminthic infections, in
CC   disinfectants and as food preservatives.
XX
SQ   Sequence 500 BP; 110 A; 144 C; 115 G; 131 T; 0 other;
```

US-10-045-180A-5 (1-44) x AAC66924 (1-500)

Alignment Scores:

Pred. No.:	0.00788	Length: 500
Score:	82.00	Matches: 19
Percent Similarity:	59.09%	Conservative: 7
Best Local Similarity:	43.18%	Mismatches: 18
Query Match:	22	Indels: 0
		Gaps: 0

```
Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlyHisGlnProProAlaAspAsp 20
         ||||:|||||||||||||:::||||||:|||||||||||||||
Db  155 GAGGCACGTCAGGCAAGAGCTGATGAAGCTCCGCCCAGCAGCCTGGAACAGATGAT 214

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
         ||||::  :::|||::  :::|||
Db  215 CAGGGAAATGGCTCATTCCTTTACACGGCCTGAAAACGCCCTCTTCCACTTTCAGAGTCA 274

Qy   41 ThrLysGlyLeu 44
         ::|||||||||
Db  275 GCGAAAGGCTTG 286
```

```
RESULT 14
AAD51768
ID   AAD51768 standard; DNA; 500 BP.
XX
AC   AAD51768;
XX
DT   16-APR-2003 (first entry)
XX
DE   Macaca mulatta RTD3 DNA.
XX
OS   Macaca mulatta.
XX
KW   Retrocyclin; infection; sexually transmitted disease; gene therapy; HIV;
KW   human immunodeficiency virus; bacterial vaginosis; ophthalmic infection;
KW   antibiotic modelling; antimicrobial; rhesus monkey; theta defensin 1A;
KW   RTD3; gene; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             95..325
FT                   /*tag= a
FT                   /product= "RTD3 protein"
FT   sig_peptide     95..154
FT                   /*tag= b
FT   mat_peptide     155..322
FT                   /*tag= c
FT                   /product= "Mature RTD3 protein"
XX
```

PN   WO20028401-A1.
XX
PD   31-OCT-2002.
XX
PF   18-APR-2002; 2002WO-US12353.
XX
PR   18-APR-2001; 2001US-284855P.
XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Lehrer RI, Waring AJ, Cole AM, Hong TB;
XX
DR   WPI; 2003-103387/09.
DR   P-PSDB; AAB38812.
XX
PT   New isolated retrocyclin peptide, useful for preventing retroviral
     infections in cells susceptible to bacterial or viral infections, or
     treating patients having the infections, such as HIV, sexually
     transmitted diseases, vaginosis -
XX
PS   Disclosure; Page 52; 72pp; English.
XX
PT   The invention relates to novel retrocyclin peptides. Peptides and
     methods of the invention are useful for preventing retroviral infections
     in cells susceptible to bacterial or viral infections, or treating
     patients having infections such as HIV (human immunodeficiency virus),
     sexually transmitted diseases, bacterial vaginosis or ophthalmic
     infections. The retrocyclin-mediated killing is useful for modelling
     and screening novel antibiotics. The invention is also useful in gene
     therapy. The present sequence is rhesus monkey theta defensin 1A
     precursor, RTD3 DNA. This sequence is used in the exemplification of
     the invention.
XX
SQ   Sequence 500 BP; 110 A; 144 C; 115 G; 131 T; 0 other;

Alignment Scores:
Pred. No.:   0.00788
Score: 82.00   length: 500
Percent Similarity: 59.09%   Matches: 19
Best Local Similarity: 43.18%   Conservative: 7
Query Match: 35.19%   Mismatches: 18
DB: 25   Indels: 0
                                               Gaps: 0

US-10-045-180A-5 (1-44) x AAD51768 (1-500)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
        ||| ||| | | | || || || | |
Db 155 GAGGCACGTCAGGCAAGAGCTGATGAAGCTGCCGCCCAGTCACCAGCCTGGGAACAGATGAT 214

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
        | |:| :: |||::| | ::::| ||| |||
Db 215 CAGGGAATGGCTCATTCCTTTACATGGGCCTGAAAACGCCGCTCTTCCACTTTCAGATGTCA 274

Qy   41 ThrLysGlyLeu 44
        |||||||||||
Db 275 GCGGAAAGGCTTG 286

RESULT 15
ID   AAQ53216 standard; cDNA; 424 BP.
XX
AC   AAQ53216;
XX
DT   25-MAR-2003 (updated)
DT   17-JUN-1994 (first entry)
XX
DE   Sequence encoding gastrointestinal defensin (GID) peptide called
     human defensin 5.
XX
KW   Gastrointestinal defensin peptide; GID; pharmaceutical; Paneth
KW   cell; antimicrobial; anti-inflammatory; diagnosis;
KW   contact disinfectant; ss.
XX

OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   10..294
FT   /*tag= a
XX
PN   WO9324513-A1.
XX
PD   09-DEC-1993.
XX
PF   18-MAY-1993; 93WO-US04740.
XX
PR   22-MAY-1992; 92US-0888232.
XX
PA   (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.
XX
PI   Bevins CL, Jones DE;
XX
DR   WPI; 1993-405719/50.
DR   P-PSDB; AAR44818.
XX
PT   Gastrointestinal defensin peptide(s) - useful as antimicrobial
     and anti-inflammatory agents and for detecting gastrointestinal
     disorders.
XX
PS   Claim 3; page 45-46; 97pp; English.
XX
PT   A probe (D5' oligo) having the sequence given in AAQ53219 (claimed)
     was used to screen a population of clones to identify novel defensin
     defensin peptides. Pref. the libraries were human genomic and cDNA
     libraries. Hybridisation and partial sequence analysis of the
     previously reported preprodefensin shows significant similarity.
     The deduced AA sequence of defensin 6 cDNA has features similar to
     defensin 5 and the previously reported preprodefensin.
     identified clones contained previously characterised myeloid derived
     defensin sequences as well as new defensin related sequences. Two
     clones expressing new defensin related sequences were extensively
     characterised and found to contain genes selectively expressed in
     Paneth cells of the small intestine. These Paneth cell-derived
     defensins are designated human defensin 5 and human defensin 6 and
     are referred to as gastrointestinal defensin peptides. Comparison
     of the deduced AA sequences of the defensin 5 cDNA with the
     previously reported preprodefensin.
     (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 424 BP; 110 A; 118 C; 95 G; 101 T; 0 other;

Alignment Scores:
Pred. No.:   0.0127
Score: 80.00   length: 424
Percent Similarity: 60.00%   Matches: 18
Best Local Similarity: 45.00%   Conservative: 6
Query Match: 34.33%   Mismatches: 16
DB: 14   Indels: 0
                                               Gaps: 0

US-10-045-180A-5 (1-44) x AAQ53216 (1-424)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
        ||| | | ||| |
Db   67 GAGTCACTCCAGGAAAGAGACTGATGAGGCTACAACCCAGAAGAAGAGTCTGGGGAGACAAC 126

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
        | |||:   :::| | :: |:| ||::
Db 127 CAGGACCTTGCTATCTCCTTTGCAGGAATGGACTCTCGCTCTTAGAACCTCAGGTTCT 186

Search completed: December 17, 2003, 15:06:03
Job time : 166 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: December 17, 2003, 15:06:18 ; Search time 152.128 Seconds
                                          (without alignments)
                                          963.776 Million cell updates/sec

Title: US-10-045-180A-5
Perfect Score: 233
Sequence: 1 EPLQARAHEMPAQKQPPADD..........IYFSGDDSCSLQVPGSTKGL 44

Scoring table: BLOSUM62
              Xgapop 10.0 , Xgapext 0.5
              Xgapop 10.0 , Ygapext 0.5
              Fgapop 6.0 , Fgapext 7.0
              Delop   6.0 , Delext 7.0

Searched: 2221978 seqs, 166101734 residues

Total number of hits satisfying chosen parameters: 4423956

Post-processing: Minimum Match 0%
              Maximum Match 10%
              listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145026_12635/app_query.fasta_1.860
-DB=Published_Applications_NA - QFMT=fastaP -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCt=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10045180 @CGN_1_1.174_@runat_17122003_145026_12635
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Published_Applications_NA.*
              1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
              2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
              3: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
              4: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
              5: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
              6: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
              7: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
              8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
              9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
             10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
             11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
             12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
             13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
             14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
             15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
             16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
             17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
             18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Match Length	DB	ID	Description

1	233	100.0	453	14	US-10-013-770-2	Sequence 2, Appli
2	233	100.0	453	14	US-10-045-180A-2	Sequence 2, Appli
3	209	89.7	4415	14	US-10-013-770-1	Sequence 1, Appli
4	209	89.7	4415	14	US-10-045-180A-1	Sequence 1, Appli
5	114	48.9	542	14	US-10-045-180A-8	Sequence 8, Appli
6	114	48.9	860	14	US-10-044-090-473	Sequence 473, App
7	104	44.6	496	13	US-10-141-645-11	Sequence 11, Appl
8	104	44.6	1348	15	US-10-125-237-78	Sequence 78, Appl
9	104	44.6	1348	15	US-10-105-891-78	Sequence 78, Appl
10	100	42.9	654	14	US-10-027-632-30054	Sequence 30054, A
11	100	42.9	592	14	US-10-027-632-30053	Sequence 30053, A
12	100	42.9	592	14	US-10-027-632-30053	Sequence 30053, A
13	100	42.9	592	14	US-10-045-180A-7	Sequence 7, Appli
14	100	42.9	741	14	US-10-141-645-120	Sequence 120, App
15	95	40.8	741	13	US-10-141-645-122	Sequence 122, App
16	95	40.8	741	13	US-10-141-645-124	Sequence 124, App
17	95	40.8	741	14	US-10-027-632-141530	Sequence 141530,
18	95	40.8	741	14	US-10-027-632-141528	Sequence 141528,
19	95	40.8	741	14	US-10-027-632-141530	Sequence 141530,
20	95	40.8	4295	14	US-10-027-632-141529	Sequence 141529,
21	90	38.6	437	13	US-10-141-645-124	Sequence 124, App
22	90	38.6	437	13	US-10-141-645-122	Sequence 122, App
23	90	38.6	437	13	US-10-141-645-120	Sequence 120, App
24	86	36.9	243	13	US-10-313-994-28	Sequence 28, Appl
25	83	35.6	495	13	US-10-141-645-16	Sequence 16, Appl
26	83	35.6	495	13	US-10-313-994-15	Sequence 15, Appl
27	82	35.2	500	13	US-10-141-645-14	Sequence 14, Appl
28	82	35.2	500	13	US-10-313-994-13	Sequence 13, Appl
29	80	34.3	2553	13	US-10-313-994-25	Sequence 25, Appl
30	78.5	33.7	337	11	US-09-918-995-18198	Sequence 18198, A
31	78.5	33.7	337	11	US-09-918-995-19070	Sequence 19070, A
32	78.5	33.7	353	11	US-09-918-995-18323	Sequence 18323, A
33	78.5	33.7	555	13	US-10-252-157-223	Sequence 223, App
34	78.5	33.7	726	10	US-09-816-828-16	Sequence 16, Appl
35	76	32.6	2525	13	US-09-313-994-24	Sequence 24, Appl
36	69	29.6	412	10	US-09-736-457-1047	Sequence 1047, Ap
37	69	29.6	412	10	US-09-902-941-1047	Sequence 1047, Ap
38	69	29.6	412	10	US-09-849-626-1047	Sequence 1047, Ap
39	69	29.6	412	13	US-10-113-872-1047	Sequence 1047, Ap
40	69	29.6	412	15	US-10-017-754-1047	Sequence 1047, Ap
41	69	29.4	597	9	US-08-864-761-15342	Sequence 15342, A
42	68.5	29.4	598	13	US-10-027-632-48280	Sequence 48280, A
43	68.5	29.4	598	13	US-10-027-632-48281	Sequence 48281, A
44	68.5	29.4	598	13	US-10-027-632-48282	Sequence 48282, A
45	68.5	29.4	598	13	US-10-027-632-48283	Sequence 48283, A

ALIGNMENTS

RESULT 1
US-10-013-770-2
Sequence 2, Application US/10013770
Publication No. US20020115151A1
GENERAL INFORMATION:
   APPLICANT: GENSET SA
   TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
                       COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
                       THERAPEUTIC APPLICATIONS
   NUMBER OF SEQUENCES: 6
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: Knobbe, Martens, Olson & Bear
      STREET: 550 West C Street
      STREET: 550 West C Street
      CITY: San Diego
      STATE: California
      COUNTRY: USA
      ZIP: 92101
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/10/013,770

FILING DATE: 10-Dec-2001
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 09/486,580
  FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
  NAME: Hart, Daniel
  REGISTRATION NUMBER: 40,637
  REFERENCE/DOCKET NUMBER: GENSET.064C1
INFORMATION FOR SEQ ID NO: 2:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 453 BASE PAIRS
    TYPE: NUCLEOTIDE
    STRANDEDNESS: DOUBLE
    TOPOLOGY: LINEAR
  MOLECULE TYPE: cDNA
  ORIGINAL SOURCE:
    ORGANISM: Homo sapiens
  SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-013-770-2

Alignment Scores:
Pred. No.:
Score: 9.49e-29   Length: 453
Percent Similarity: 233.00   Matches: 44
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 14   Indels: 0
                                          Gaps: 0

US-10-045-180A-5 (1-44) x US-10-013-770-2 (1-453)

```
Qy    1  GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnIleLysGlnProProAlaAsbAsp   20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109  GAGCCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGAC    168

Qy   21  GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer    40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  169  CAGGATGTGGTCATTTACTTTCAGGAGATGACAGCTGCTTCTTCAGGTTCCAGGCTCA      228

Qy   41  ThrLysGlyLeu   44
         ||||||||
Db  229  ACAAAGGGCTTG   240
```

RESULT 2
US-10-045-180A-2
Sequence 2, Application US/10045180A
Publication No. US20020182703A1
GENERAL INFORMATION:
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Compositi
TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeutic
FILE REFERENCE: GEN-100D1
CURRENT APPLICATION NUMBER: US/10/045,180A
CURRENT FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: US 09/486,580
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: PCT/FR98/01864
PRIOR FILING DATE: 1998-08-28
PRIOR APPLICATION NUMBER: FR 97/10823
PRIOR FILING DATE: 1997-08-29
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 453
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (52)..(336)
OTHER INFORMATION: Def-X coding sequence
US-10-045-180A-2

Alignment Scores:

RESULT 3
US-10-013-770-1
Sequence 1, Application US/10013770
Publication No. US20020115151A1
GENERAL INFORMATION:
APPLICANT: GENSET SA
TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
THERAPEUTIC APPLICATIONS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 550 West C Street
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/013,770
FILING DATE: 10-Dec-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/486,580
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hart, Daniel
REGISTRATION NUMBER: 40,637
REFERENCE/DOCKET NUMBER: GENSET.064C1
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4415 BASE PAIRS
TYPE: NUCLEOTIDE
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Exon 1
LOCATION: 1836..1874
FEATURE:
NAME/KEY: Exon 2
LOCATION: 3394..3577
FEATURE:
NAME/KEY: Exon 3
LOCATION: 4161..4380
FEATURE:
NAME/KEY: start CDS

Pred. No.:
Score: 9.49e-29   Length: 453
Percent Similarity: 233.00   Matches: 44
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 14   Indels: 0
                                          Gaps: 0

US-10-045-180A-5 (1-44) x US-10-045-180A-2 (1-453)

```
Qy    1  GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnIleLysGlnProProAlaAsbAsp   20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109  GAGCCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGAC    168

Qy   21  GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer    40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  169  CAGGATGTGGTCATTTACTTTCAGGAGATGACAGCTGCTTCTTCAGGTTCCAGGCTCA      228

Qy   41  ThrLysGlyLeu   44
         ||||||||
Db  229  ACAAAGGGCTTG   240
```

```
;             FEATURE:
;                 LOCATION: 3406..3408
;             NAME/KEY: stop CDS
;                 LOCATION: 4276..4278
;             FEATURE:
;                 NAME/KEY: polyadenylation site
;                 LOCATION: 4374..4379
US-10-013-770-1
;     SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Alignment Scores:
   Pred. No.:
   Score:                    1.19e-23
   Percent Similarity:       209.00            Length:           4415
   Best Local Similarity:    100.00%           Matches:          39
   Query Match:              100.00%           Conservative:     0
   DB:                       89.70%            Mismatches:       0
                             14                Indels:           0
                                               Gaps:             0

US-10-045-180A-5 (1-44) x US-10-013-770-1 (1-4415)

Qy     1 GluProLeuGlnAlaArgAlaHisGlnMetProAlaGlnLysGlnProProAlaAspAsp 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3463 GAGCCCGCTTCCAGGCAAGAGCTTCATGAGAATGCCAGCCCAGAGGCAGCTTCCAGCAGAGATGAC 3522

Qy    21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly 39
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3523 CAGGATGAGTGGTCATTTACTTTTCAGAGAGATGACAGTCTGCTCTCTTCTTCAGGTTCCAGGT 3579

RESULT 4
US-10-045-180A-1
;   Sequence 1, Application US/10045180A
;   Publication No. US20020182703A1
;   GENERAL INFORMATION:
;       APPLICANT: Bougueleret, Lydie
;       APPLICANT: Chumakov, Ilya
;       TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Compositi
;       TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeutic
;       FILE REFERENCE: GEN-100D1
;       CURRENT APPLICATION NUMBER: US/10/045,180A
;       CURRENT FILING DATE: 2001-10-18
;       PRIOR APPLICATION NUMBER: US 09/486,580
;       PRIOR FILING DATE: 2000-02-25
;       PRIOR APPLICATION NUMBER: PCT/FR98/01864
;       PRIOR FILING DATE: 1998-08-28
;       PRIOR APPLICATION NUMBER: FR 97/10823
;       PRIOR FILING DATE: 1997-08-29
;       NUMBER OF SEQ ID NOS: 12
;       SOFTWARE: Patentin version 3.1
;
;   SEQ ID NO 1
;       LENGTH: 4415
;       TYPE: DNA
;       ORGANISM: Homo sapiens
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (1)..(4415)
;           OTHER INFORMATION: Def-X genomic sequence
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (85)..(85)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (143)..(143)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (670)..(670)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (970)..(970)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
```

```
;           NAME/KEY: misc_feature
;           LOCATION: (1111)..(1111)
;           OTHER INFORMATION:
;       FEATURE:
;           NAME/KEY: TATA_signal
;           LOCATION: (1758)..(1767)
;           OTHER INFORMATION: Exon 1
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (1150)..(1150)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (1780)..(1780)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
;           NAME/KEY: CAAT_signal
;           LOCATION: (1711)..(1714)
;           OTHER INFORMATION:
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (1836)..(1874)
;           OTHER INFORMATION: Exon 1
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (1974)..(1974)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (2117)..(2117)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (1875)..(1880)
;           OTHER INFORMATION: splice donor site
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (2133)..(2133)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (2155)..(2335)
;           OTHER INFORMATION: Alu insertion
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (2186)..(2186)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (2191)..(2191)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (2367)..(2367)
;           OTHER INFORMATION: n = a, c, g, or t.
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (2710)..(2780)
;           OTHER INFORMATION: L1 fragment insertion
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (3391)..(3393)
;           OTHER INFORMATION: splice acceptor site
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (3394)..(3577)
;           OTHER INFORMATION: Exon 2
;       FEATURE:
;           NAME/KEY: misc_feature
;           LOCATION: (3406)..(3408)
;           OTHER INFORMATION: Translation initiation codon (ATG)
;       FEATURE:
;           NAME/KEY: misc_feature
```

; LOCATION: (3578)..(3583)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: splice donor site
; LOCATION: (4123)
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: n = a, c, g, or t.
; LOCATION: (4123)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4161)..(4163)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: splice acceptor site
; LOCATION: (4164)..(4379)
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Exon 3
; LOCATION: (4274)..(4276)
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Translation termination codon (TAA)
; LOCATION: (4374)..(4379)
; FEATURE:
; NAME/KEY: polyA signal
; OTHER INFORMATION:
US-10-045-180A-1

Alignment Scores:

Pred. No.:   Length: 4415
Score: 1.19e-23   Matches: 39
Percent Similarity: 209.00   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 89.70%   Gaps: 0
                             14

US-10-045-180A-5 (1-44) x US-10-045-180A-1 (1-4415)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3463 GAGCCGCTCCAGGCAAGAGCCCATGAGAATGCCAGCCCCAGAAGCAGCTCCCAGAGATGAC 3522

Qy   21 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly 39
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3523 CAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTCCTCTCTTCAGGTTCCAGGT 3579

RESULT 5
US-10-045-180A-8
; Sequence 8, Application US/10045180A
; Publication No. US20020182703A1
; GENERAL INFORMATION:
; APPLICANT: Bouguelaret, Lydie
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Composit
; FILE REFERENCE: GEN-100D1
; CURRENT APPLICATION NUMBER: US/10/045,180A
; CURRENT FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 09/486,580
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: PCT/FR98/01864
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: FR 97/10823
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8
; LENGTH: 542
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52)..(345)
; OTHER INFORMATION: Def-4 (HNP-4) coding sequence

Alignment Scores:

Pred. No.:   Length: 542
Score: 3.26e-09   Matches: 23
Percent Similarity: 114.00   Conservative: 7
Best Local Similarity: 69.77%   Mismatches: 13
Query Match: 53.49%   Indels: 0
DB: 48.93%   Gaps: 0
                             14

US-10-045-180A-5 (1-44) x US-10-045-180A-8 (1-542)

Qy   2 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 21
            ||||||||| ||||| ||| ||||||||||||| ||:||||||||||||||||||
Db   112 CCACTCCAGGCAAGAGCCCATGAGAGCTCCAGGACAGGAGCAGGAGAGGAGGCCCAGAGA 171

Qy   22 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 41
            |||:|| |||||||||||||||||:::|||||||||||||||||||||
Db   172 GACATATCTATTTCCTTTGCATGGGATAAAAAGCTTCGTGCTCTTCAGGTTTCAGGCTCAACA 231

Qy   42 LysGlyLeu 44
            :::|||:||:
Db   232 AGGGGCATG 240

RESULT 6
US-10-044-090-473
; Sequence 473, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 473
; LENGTH: 860
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1670142CB1

Alignment Scores:

Pred. No.:   Length: 860
Score: 5.65e-09   Matches: 23
Percent Similarity: 114.00   Conservative: 7
Best Local Similarity: 69.77%   Mismatches: 13
Query Match: 53.49%   Indels: 0
DB: 48.93%   Gaps: 0
                             14

US-10-045-180A-5 (1-44) x US-10-044-090-473 (1-860)

Qy   2 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 21
            ||||||||| ||||| ||| ||||||||||||| ||:||||||||||||||||||
Db   136 CCACTCCAGGCAAGAGGTGATGAGGCTCCAGGACGTGGGCGCCAGAAGACCAG 195

Qy   22 AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySerThr 41
            |||:||| |||||||||||||||||:::|||||||||||||||||||||
Db   196 GACATATCTATTTCCTTTGCATGGGATAAAAAGCTTCGTGCTCTTCAGGTTTCAGGCTCAACA 255

Qy   42 LysGlyLeu 44
            :::|||:||:
Db   256 AGGGGCATG 264

RESULT 7
US-10-141-645-11
; Sequence 11, Application US/10141645
; Publication No. US20030144184A1
; GENERAL INFORMATION:
; APPLICANT: Robert Lehrer
; APPLICANT: Alan Waring
; APPLICANT: Alexander Cole
; APPLICANT: Teresa Hong

TITLE OF INVENTION: Retrocyclins - Antiviral and
TITLE OF INVENTION: Antimicrobial Peptides
FILE REFERENCE: UCLA-001CIP
CURRENT APPLICATION NUMBER: US/10/141,645
CURRENT FILING DATE: 2002-05-06
PRIOR APPLICATION NUMBER: 60/284,855
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: Unassigned
PRIOR FILING DATE: 2002-04-18
NUMBER OF SEQ ID NOS: 125
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 11
LENGTH: 496
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (124)...(304)
OTHER INFORMATION: retrocyclin
US-10-141-645-11

US-10-045-180A-5 (1-44) x US-10-141-645-11 (1-496)

Alignment Scores:
Pred. No.:   1.26e-07
Score: 104.00
Percent Similarity: 68.18%
Best Local Similarity: 50.00%
Query Match: 44.64%
DB: 13

Length: 496
Matches: 22
Conservative: 8
Mismatches: 14
Indels: 0
Gaps: 0

```
Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
          ||||||||||||::|||||:||:::||||||||||||:||:||||||||||||||
Db  136 GAGCCACTTCAGGCAAGCTGCAGAGCTGCAGCCCAGGAGCAGCCTGGAGCAGATGAT 195

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
          |||:|||::|:::|||:
Db  196 CAGGAAAATGGCTCATGCCTTTACATGGCATGAAAGTCCGGCTCTTCCGCTTTCAGACTCA 255

Qy   41 ThrLysGlyLeu 44
          :::||||||
Db  256 GCGGAGAGCTTG 267
```

RESULT 8
US-10-125-237-78
Sequence 78, Application US/10125237
Publication No. US2003022329A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Zhou, Ping
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyan
APPLICANT: Zhao, Qing A.
APPLICANT: Xue, Aidong J.
APPLICANT: Zhang, Jie
APPLICANT: Wehrman, Tom
APPLICANT: Wang, Jian-Rui
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: Polypeptides
TITLE OF INVENTION: No. US2003022329A1e1 Nucleic Acids and
FILE REFERENCE: 791CIP2ADIV
CURRENT APPLICATION NUMBER: US/10/125,237
CURRENT FILING DATE: 2002-04-17
PRIOR APPLICATION NUMBER: 09/668,317
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 09/552,929
PRIOR FILING DATE: 2000-04-18
NUMBER OF SEQ ID NOS: 91
SOFTWARE: pt_FL_genes Version 2.0
SEQ ID NO 78
LENGTH: 1348
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1143)
US-10-125-237-78

US-10-045-180A-5 (1-44) x US-10-125-237-78 (1-1348)

Alignment Scores:
Pred. No.:   4.15e-07
Score: 104.00
Percent Similarity: 68.29%
Best Local Similarity: 56.10%
Query Match: 44.64%
DB: 15

Length: 1348
Matches: 23
Conservative: 5
Mismatches: 13
Indels: 0
Gaps: 0

```
Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
          |||||||||||||:|||   |||:::|||||||||||||:||:||||||||||||
Db  421 GAGCCACTTCAGGCAATTGCTGATGAGGGCTACAGCCCAGGAGCAGCCTGGAGCAGATGAT 480

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
          |||:|||||||:::  |||||||||||||||:::
Db  481 CAGGAAGTGGTTGGATTCCTTTGCATGGCATGAAAGAGCTCCCTCTTCAGGTTTCAGGGAAG 540

Qy   41 Thr 41
          :::
Db  541 TCC 543
```

RESULT 9
US-10-105-891-78
Sequence 78, Application US/10105891
Publication No. US20030073099A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Zhou, Ping
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyan
APPLICANT: Zhao, Qing A.
APPLICANT: Xue, Aidong J.
APPLICANT: Zhang, Jie
APPLICANT: Wehrman, Tom
APPLICANT: Wang, Jian-Rui
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: Polypeptides
TITLE OF INVENTION: No. US20030073099A1e1 Nucleic Acids and
FILE REFERENCE: 791CIP2A
CURRENT APPLICATION NUMBER: US/10/105,891
CURRENT FILING DATE: 2002-03-25
PRIOR APPLICATION NUMBER: 09/668,317
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 09/552,929
PRIOR FILING DATE: 2000-04-18
NUMBER OF SEQ ID NOS: 91
SOFTWARE: pt_FL_genes Version 2.0
SEQ ID NO 78
LENGTH: 1348
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1143)
US-10-105-891-78

US-10-045-180A-5 (1-44) x US-10-105-891-78 (1-1348)

RESULT 10
US-10-027-632-30053
Sequence 30053, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30053
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(592)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-30053

Alignment Scores:
Pred. No.:
Score: 7.02e-07
Percent Similarity: 100.00
Best Local Similarity: 68.29%
Query Match: 51.22%
DB: 42.92%

Length: 592
Matches: 21
Conservative: 7
Mismatches: 13
Indels: 0
Gaps: 0

US-10-045-180A-5 (1-44) x US-10-027-632-30053 (1-592)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnHisGlnProProAlaAspAsp 20
Db 215 GAGCCACTTCAGGCAAGAGCTGATGAAGCTGCAGCCCCAGGAGCAGCCTGAGCAGATGAT 274
Qy 21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
Db 275 CAGGAAATGGCTCATGCCTTTACATGGCATGGAAAGAGTGCCGCCTCTTCCGGCTTTCAGGTGAG 334
Qy 41 Thr 41
        |||
Db 335 ACA 337

RESULT 11
US-10-027-632-30053

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnHisGlnProProAlaAspAsp 20
Db 421 GAGCCACTCCAGGCAARTTGCTGATGAGGCTTACAGCCCAGGAGCAGCCTGGAGCAGATGAT 480
Qy 21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
Db 481 CAGGAAGTGGTTGATTCCTTTGCATGGGATGGAAAGAGCTCTTCTTCCGGTTTCAGGTGAAG 540
Qy 41 Thr 41
        :::
Db 541 TCC 543

US-10-027-632-30054
Sequence 30054, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30054
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(592)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-30054

Alignment Scores:
Pred. No.:
Score: 7.02e-07
Percent Similarity: 100.00
Best Local Similarity: 68.29%
Query Match: 51.22%
DB: 42.92%

Length: 592
Matches: 21
Conservative: 7
Mismatches: 13
Indels: 0
Gaps: 0

US-10-045-180A-5 (1-44) x US-10-027-632-30054 (1-592)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnHisGlnProProAlaAspAsp 20
Db 215 GAGCCACTTCAGGCAAGAGCTGATGAAGCTGCAGCCCCAGGAGCAGCCTGAGCAGATGAT 274
Qy 21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
Db 275 CAGGAAATGGCTCATGCCTTTACATGGCATGGAAAGAGTGCCGCCTCTTCCGGCTTTCAGGTGAG 334
Qy 41 Thr 41
        |||
Db 335 ACA 337

RESULT 12
US-10-027-632-30053
Sequence 30053, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483

```
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 30053
;  LENGTH: 592
;  TYPE: DNA
;  ORGANISM: Human
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(592)
;  OTHER INFORMATION: n = A,T,C or G
US-10-027-632-30053

Alignment Scores:
Pred. No.:
Score:                  7.02e-07
Percent Similarity:     68.29%
Best Local Similarity:  51.22%
Query Match:            42.92%
DB:

                        Length:        592
                        Matches:       21
                        Conservative:  7
                        Mismatches:    13
                        Indels:        0
                        Gaps:          0

US-10-045-180A-5 (1-44) x US-10-027-632-30053 (1-592)

Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
Db  215 GAGCCACTTCAGGCAAGAGCTGATGAAGCTGCAGCCAAGGAGCAGCCTGGAGCAGGATGAT 274

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
Db  275 CAGGAAAATGGCTCATGCCTTTACATGGCATGAAAAGTGCCGCTCTTCGGCTTTCAGGTGAG 334

Qy   41 Thr 41
Db  335 ACA 337


RESULT 13
US-10-027-632-30054
;  Sequence 30054, Application US/10027632
;  GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION: Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE: 2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 30054
;  LENGTH: 592
;  TYPE: DNA
;  ORGANISM: Human
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(592)
;  OTHER INFORMATION: n = A,T,C or G
US-10-027-632-30054

Alignment Scores:
Pred. No.:
Score:                  7.02e-07
Percent Similarity:     68.29%
Best Local Similarity:  51.22%
Query Match:            42.92%
DB:

                        Length:        592
                        Matches:       21
                        Conservative:  7
                        Mismatches:    13
                        Indels:        0
                        Gaps:          0

US-10-045-180A-5 (1-44) x US-10-027-632-30054 (1-592)

Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
Db  215 GAGCCACTTCAGGCAAGAGCTGATGAAGCTGCAGCCAAGGAGCAGCCTGGAGCAGGATGAT 274

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
Db  275 CAGGAAAATGGCTCATGCCTTTACATGGCATGAAAAGTGCCGCTCTTCGGCTTTCAGGTGAG 334

Qy   41 Thr 41
Db  335 ACA 337


RESULT 14
US-10-027-632-141529
;  Sequence 141529, Application US/10027632
;  Publication No. US20030204075A9
;  GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION: Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE: 2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 141529
;  LENGTH: 654
;  TYPE: DNA
;  ORGANISM: Human
US-10-027-632-141529

Alignment Scores:
Pred. No.:
Score:                  5.18e-06
Percent Similarity:     95.00%
Best Local Similarity:  65.79%
Query Match:            52.63%
DB:

                        Length:        654
                        Matches:       20
                        Conservative:  5
                        Mismatches:    13
                        Indels:        0
                        Gaps:          0

US-10-045-180A-5 (1-44) x US-10-027-632-141529 (1-654)

Qy    2 ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 21
```

```
Db   96  CCACTCCAGGCAAGAGGGTGATGAGGCTCCAGGCCACCAGCACCGTGGGCCAGAAGACCAG 155

Qy   22  AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly 39
         ||||:::     |||    |||   :::||||||||||| |||
Db  156  GACATATCTATTCCTTTGCATGGGATAAAAGCTCTGCTCTTCAGGTTTCAGGT 209
```

RESULT 15
US-10-027-632-141529
Sequence 141529, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141529
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-141529

Alignment Scores:
Pred. No.:
Score: 5.18e-06   Length: 654
Percent Similarity: 95.00   Matches: 20
Best local Similarity: 65.79%   Conservative: 5
Query Match: 52.63%   Mismatches: 13
DB: 40.77%   Indels: 0
                      14   Gaps: 0

US-10-045-180A-5 (1-44) x US-10-027-632-141529 (1-654)

```
Qy    2  ProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAspGln 21
         |||||||||||||||||||  |||    |||:::|||||||
Db   96  CCACTCCAGGCAAGAGGGTGATGAGGCTCCAGGCCACCAGCACCGTGGGCCAGAAGACCAG 155

Qy   22  AspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly 39
         ||:::     |||    |||   :::|||||||||||| |||
Db  156  GACATATCTATTCCTTTGCATGGGATAAAAGCTCTGCTCTTCAGGTTTCAGGT 209
```

Search completed: December 17, 2003, 21:36:01
Job time : 154.128 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: December 17, 2003, 14:50:33 ; Search time 2308.83 Seconds
(without alignments)
632.011 Million cell updates/sec

Title:
Perfect score: 233
Sequence: 1 BPLQARAHEMPAQKQPPADD..........IYFSGDDSCSLQVPGSTKGL 44

Searched: US-10-045-180A-5

Total number of hits satisfying chosen parameters: 3336368 seqs, 1658189874 residues

66727376

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line Parameters:
-MODEL=frame+_p2n.model
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145025_12578/app_query.fasta_1.860
-DB=Pending_Patents_NA_Main -QFMT=fastp -SUFFIX=rnpm -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10045180 @CGN_1_9409 @runat_17122003_145025_12578 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.old:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US1009C_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*

102: /cgm2_6/ptodata/2/pna/US6047_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a,
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	233	100.0	285	47	US-10-170-235-6564	Sequence 6564, Ap
2	233	100.0	453	44	US-10-013-770-2	Sequence 2, Appli
3	233	100.0	453	44	US-10-045-180A-2	Sequence 2, Appli
4	209	89.7	4415	44	US-10-013-770-1	Sequence 1, Appli
5	209	89.7	4415	44	US-10-045-180A-1	Sequence 1, Appli
6	200	85.8	601	40	US-09-947-907-9094	Sequence 9094, Ap
7	200	85.8	610	70	US-60-160-203-825	Sequence 825, App
8	200	85.8	610	70	US-60-169-840-3181	Sequence 3181, Ap
9	200	85.8	696	70	US-60-160-203-156	Sequence 156, App
10	200	85.8	111999	75	US-60-212-664-312	Sequence 312, App
11	187	80.3	345	71	US-60-171-481-272	Sequence 272, APP
12	187	80.3	345	71	US-60-171-481-777	Sequence 777, APP
13	120	51.5	338	19	US-09-321-214-10452	Sequence 10452, A
14	120	51.5	338	22	US-09-516-335-10452	Sequence 10452, A
15	120	51.5	338	32	US-09-733-811-10452	Sequence 10452, A
16	120	51.5	338	32	US-09-733-811A-10452	Sequence 10452, A
17	120	51.5	338	42	US-09-975-640-10452	Sequence 10452, A
18	120	51.5	338	42	US-09-975-640A-10452	Sequence 10452, A
19	118	50.6	338	21	US-09-471-275-1730	Sequence 1730, Ap
20	118	50.6	338	21	US-09-488-725B-6871	Sequence 6871, Ap
21	114	48.9	271	16	US-09-179-473-1553	Sequence 1553, Ap
22	114	48.9	271	19	US-09-328-351-1553	Sequence 1553, Ap
23	114	48.9	271	20	US-09-359-067-26554	Sequence 26554, A
24	114	48.9	271	35	US-09-845-487-1553	Sequence 1553, Ap
25	114	48.9	271	35	US-09-845-487A-1553	Sequence 1553, Ap
26	114	48.9	304	18	US-09-271-490-2574	Sequence 2574, Ap
27	114	48.9	304	39	US-09-925-552-1874	Sequence 1874, Ap
28	114	48.9	304	44	US-10-032-354-1874	Sequence 1874, Ap
29	114	48.9	349	19	US-09-321-214-3560	Sequence 3560, Ap
30	114	48.9	349	22	US-09-516-335-3560	Sequence 3560, Ap
31	114	48.9	349	32	US-09-733-811-3560	Sequence 3560, Ap
32	114	48.9	349	32	US-09-733-811A-3560	Sequence 3560, Ap
33	114	48.9	349	42	US-09-975-640-3560	Sequence 3560, Ap
34	114	48.9	349	42	US-09-975-640A-3560	Sequence 3560, Ap
35	114	48.9	350	18	US-09-271-490-2574	Sequence 2574, Ap
36	114	48.9	350	39	US-09-925-552-2574	Sequence 2574, Ap
37	114	48.9	350	44	US-10-032-354-2574	Sequence 2574, Ap
38	114	48.9	382	17	US-09-234-611-10800	Sequence 10800, A
39	114	48.9	382	18	US-09-289-768-15603	Sequence 15603, A
40	114	48.9	382	38	US-09-904-809-10800	Sequence 10800, A
41	114	48.9	382	39	US-09-939-397-15603	Sequence 15603, A
42	114	48.9	396	18	US-09-289-768-13222	Sequence 13222, A
43	114	48.9	396	39	US-09-939-397-13222	Sequence 13222, A
44	114	48.9	521	32	US-09-726-811-171	Sequence 171, App
45	114	48.9	524	34	US-09-823-241-2013	Sequence 2013, Ap

ALIGNMENTS

RESULT 1
US-10-170-235-6564
Sequence 6564, Application US/10170235
; GENERAL INFORMATION:
; APPLICANT: VENTER, J Craig
; TITLE OF INVENTION: KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF HUMAN
; TITLE OF INVENTION: TRANSCRIPTS, FOR DETECTING EXPRESSION AND OTHER USES THEREOF
; FILE REFERENCE: CL001380
; CURRENT APPLICATION NUMBER: US/10/170,235
; CURRENT FILING DATE: 2003-03-17
; NUMBER OF SEQ ID NOS: 42514
; SEQ ID NO 6564
; LENGTH: 285

; TYPE: DNA
; ORGANISM: HUMAN
US-10-170-235-6564
Alignment Scores:
Pred. No.:   1.25e-22
Score: 233.00   Length: 285
Percent Similarity: 100.00%   Matches: 44
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 47   Mismatches: 0
DB: Indels: 0
                                              Gaps: 0

US-10-045-180A-5 (1-44) x US-10-170-235-6564 (1-285)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 58 GAGCCGCTTCCAGGCAAGAGAGCTCATGAGAATGCCAGCCCAGAAGACAGCCTCCAGCAGAGATGAC 117

Qy 21 GlnAspValValIleTyrPheSerGlyAspPheSerCysSerLeuGlnValProGlySer 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 118 CAGGATGTGGTGATTTACTTTTCAGGAGATGACAGCTGCTCTCTCTTCAGGTTCCAGGCTCA 177

Qy 41 ThrLlyGlyLeu 44
        ||||||||||||
Db 178 ACAAAGGGCTTG 189

RESULT 2
US-10-013-770-2
Sequence 2, Application US/10013770
; GENERAL INFORMATION:
; APPLICANT: GENSET SA
; TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;   COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;   THERAPEUTIC APPLICATIONS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/013,770
;   FILING DATE: 10-Dec-2001
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/486,580
;   FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 453 BASE PAIRS
;   TYPE: NUCLEOTIDE
;   STRANDEDNESS: DOUBLE
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE: cDNA
;   ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
;   SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-013-770-2
Alignment Scores:
Pred. No.:   2.15e-22
Score: 233.00   Length: 453
Percent Similarity: 100.00%   Matches: 44
                                              Conservative: 0

Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 44   Gaps: 0

US-10-045-180A-5 (1-44) x US-10-013-770-2 (1-453)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109 GAGCCGCTCCAGGCAACGAGCTCATGAGAATGCCAGCCCAGAAGCAGCCTCCAGCAGATGAC 168

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 169 CAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGCTCA  228

Qy  41 ThrLysGlyLeu  44
         ||||||||||||
Db 229 ACAAAGGGCTTG  240
```

RESULT 3
US-10-045-180A-2
; Sequence 2, Application US/10045180A
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Compositi
; TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeutic
; FILE REFERENCE: GEN-100D1
; CURRENT APPLICATION NUMBER: US/10/045,180A
; CURRENT FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 09/486,580
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: PCT/FR98/01864
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: FR 97/10823
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52)..(336)
; OTHER INFORMATION: Def-X coding sequence

US-10-045-180A-2

US-10-045-180A-5 (1-44) x US-10-045-180A-2 (1-453)

Alignment Scores:
Pred. No.:
Score: 2.15e-22   233.00   Length: 453
Percent Similarity: 100.00%   Matches: 44
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: Indels: 0
    Gaps: 0

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109 GAGCCGCTCCAGGCAACGAGCTCATGAGAATGCCAGCCCAGAAGCAGCCTCCAGCAGATGAC 168

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 169 CAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGCTCA  228

Qy  41 ThrLysGlyLeu  44
         ||||||||||||
Db 229 ACAAAGGGCTTG  240
```

RESULT 4
US-10-013-770-1
; Sequence 1, Application US/10013770
; GENERAL INFORMATION:

APPLICANT: GENSET SA
TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
THERAPEUTIC APPLICATIONS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 550 West C Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/013,770
; FILING DATE: 10-Dec-2001
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/486,580
; FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
; NAME: Hart, Daniel
; REGISTRATION NUMBER: 40,637
; REFERENCE/DOCKET NUMBER: GENSET.064C1
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
; LENGTH: 4415 BASE PAIRS
; TYPE: NUCLEOTIDE
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
FEATURE:
; NAME/KEY: Exon 1
; LOCATION: 1836..1874
FEATURE:
; NAME/KEY: Exon 2
; LOCATION: 3394..3577
FEATURE:
; NAME/KEY: Exon 3
; LOCATION: 4161..4380
FEATURE:
; NAME/KEY: start CDS
; LOCATION: 3406..3408
FEATURE:
; NAME/KEY: stop CDS
; LOCATION: 4276..4278
FEATURE:
; NAME/KEY: polyadenylation site
; LOCATION: 4374..4379
SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-10-013-770-1

Alignment Scores:
Pred. No.:
Score: 7e-18   209.00   Length: 4415
Percent Similarity: 100.00%   Matches: 39
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 89.70%   Mismatches: 0
DB: 44   Indels: 0
    Gaps: 0

US-10-045-180A-5 (1-44) x US-10-013-770-1 (1-4415)

```
Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp  20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3463 GAGCCGCTCCAGGCAACGAGCTCATGAGAATGCCAGCCCAGAAGCAGCCTCCAGCAGATGAC 3522

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly  39
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db

RESULT 5

3523 CAGGAGGTGGTCATTTACTTTCAGGAGGATGACAGCTGCTCTCTTCAGGTTCCAGGT 3579

US-10-045-180A-1
; Sequence 1, Application US/10045180A
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Compositi
; TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeutic
; FILE REFERENCE: GEN-100D1
; CURRENT APPLICATION NUMBER: US/10/045,180A
; CURRENT FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 09/486,580
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: PCT/FR98/01864
; PRIOR FILING DATE: 1998-09-28
; PRIOR APPLICATION NUMBER: FR 97/10823
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
;
SEQ ID NO 1
; LENGTH: 4415
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(4415)
; OTHER INFORMATION: Def-X genomic sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (143)..(143)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (85)..(85)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (670)..(670)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (970)..(970)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: CAAT_signal
; LOCATION: (1711)..(1714)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: TATA_signal
; LOCATION: (1758)..(1767)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1111)..(1111)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1150)..(1150)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1780)..(1780)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1836)..(1874)
; OTHER INFORMATION: Exon 1
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1875)..(1880)
; OTHER INFORMATION: splice donor site

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1974)..(1974)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2117)..(2117)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2133)..(2133)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2155)..(2335)
; OTHER INFORMATION: Alu insertion
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2186)..(2186)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2191)..(2191)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2367)..(2367)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3391)..(3393)
; OTHER INFORMATION: splice acceptor site
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3394)..(3577)
; OTHER INFORMATION: Exon 2
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2710)..(2780)
; OTHER INFORMATION: L1 fragment insertion
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3406)..(3408)
; OTHER INFORMATION: Translation initiation codon (ATG)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3578)..(3583)
; OTHER INFORMATION: splice donor site
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4123)..(4123)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4161)..(4163)
; OTHER INFORMATION: splice acceptor site
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4164)..(4379)
; OTHER INFORMATION: Exon 3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4274)..(4276)
; OTHER INFORMATION: Translation termination codon (TAA)
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: (4374)..(4379)
; OTHER INFORMATION:
US-10-045-180A-1

Alignment Scores:
Pred. No.:   7e-18   Length: 4415
Score: 209.00   Matches: 39

Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 89.70%   Indels: 0
DB: 44   Gaps: 0

US-10-045-180A-5 (1-44) x US-10-045-180A-1 (1-4415)

```
Qy    1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3463 GAGCCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCCAGAAAGCAGCCTCCAGCAGATGAC 3522

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerSerLeuGlnValProGly 39
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3523 CAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTTCTTCAGGTTCCAGGT 3579
```

RESULT 6
US-09-947-907-9094/c
; Sequence 9094, Application US/09947907
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: COLLECTION OF CODING REGION SINGLE
; TITLE OF INVENTION: NUCLEOTIDE POLYMORPHISM (cSNPs) LOCATED ON EACH OF THE HUMAN
; TITLE OF INVENTION: CHROMOSOMES, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL000896
; CURRENT APPLICATION NUMBER: US/09/947,907
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 21266
; SEQ ID NO 9094
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-947-907-9094

Alignment Scores:
Pred. No.:   Length: 601
Score: 200.00   Matches: 38
Percent Similarity: 97.44%   Conservative: 0
Best Local Similarity: 97.44%   Mismatches: 1
Query Match: 85.84%   Indels: 0
DB: 40   Gaps: 0
                            1.24e-17

US-10-045-180A-5 (1-44) x US-09-947-907-9094 (1-601)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 322 GAGCCGCTCCAGGCAAGAGCTSATGAGATGCCAGCCCAGAAAGCAGCCTCCAGCAGATGAC 263

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerSerLeuGlnValProGly 39
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 262 CAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTCTTCAGGTTCCAGGT 206
```

RESULT 7
US-60-160-203-825/c
; Sequence 825, Application US/60160203
; GENERAL INFORMATION:
; APPLICANT: BONAZZI, VIVIEN
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000116
; CURRENT APPLICATION NUMBER: US/60/160,203
; CURRENT FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 6374
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 825
; LENGTH: 610
; TYPE: DNA
; ORGANISM: HUMAN
US-60-160-203-825

Alignment Scores:
Pred. No.:   Length: 610
Score: 200.00   Matches: 38
Percent Similarity: 97.44%   Conservative: 0
Best Local Similarity: 97.44%   Mismatches: 1
Query Match: 85.84%   Indels: 0
DB: 70   Gaps: 0
                            1.26e-17

US-10-045-180A-5 (1-44) x US-60-160-203-825 (1-610)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 356 GAGCCGCTCCAGGCAAGAGCTGATGAGATGCCAGCCCAGAAAGCAGCCTCCAGCAGATGAC 297

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerSerLeuGlnValProGly 39
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296 CAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTTCTTCAGGTTCCAGGT 240
```

RESULT 8
US-60-169-840-3181/c
; Sequence 3181, Application US/60169840
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000164
; CURRENT APPLICATION NUMBER: US/60/169,840
; CURRENT FILING DATE: 1999-12-09
; NUMBER OF SEQ ID NOS: 9628
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3181
; LENGTH: 610
; TYPE: DNA
; ORGANISM: Human
US-60-169-840-3181

Alignment Scores:
Pred. No.:   Length: 610
Score: 200.00   Matches: 38
Percent Similarity: 97.44%   Conservative: 0
Best Local Similarity: 97.44%   Mismatches: 1
Query Match: 85.84%   Indels: 0
DB: 70   Gaps: 0
                            1.26e-17

US-10-045-180A-5 (1-44) x US-60-169-840-3181 (1-610)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 356 GAGCCGCTCCAGGCAAGAGCTGATGAGATGCCAGCCCAGAAAGCAGCCTCCAGCAGATGAC 297

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerSerLeuGlnValProGly 39
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296 CAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTTCTTCAGGTTCCAGGT 240
```

RESULT 9
US-60-160-203-156
; Sequence 156, Application US/60160203
; GENERAL INFORMATION:
; APPLICANT: BONAZZI, VIVIEN
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000116
; CURRENT APPLICATION NUMBER: US/60/160,203
; CURRENT FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 6374
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 156

; LENGTH: 696
; TYPE: DNA
; ORGANISM: HUMAN
US-60-160-203-156

Alignment Scores:

Pred. No.:	1.47e-17	
Score:	200.00	Length:	696
Percent Similarity:	97.44%	Matches:	38
		Conservative:	0
Best Local Similarity:	97.44%	Mismatches:	1
Query Match:	85.84%	Indels:	0
		Gaps:	0

US-10-045-180A-5 (1-44) x US-60-160-203-156 (1-696)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 415 GAGCCCGCTCCAGGCAAGAGCTGATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGAC 474

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerSerLeuGlnValProGly 39
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 475 CAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTTCTTCAGGTTCCAGGT 531
```

RESULT 10
US-60-212-664-312
; Sequence 312, Application US/60212664
; GENERAL INFORMATION:
APPLICANT: Ladunga, Steve
APPLICANT: Spier, Gene
APPLICANT: Greenberg, Simon
APPLICANT: Rabkin, Steven
APPLICANT: Wang, Yu
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS, AND
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: CL000687
CURRENT APPLICATION NUMBER: US/60/212,664
CURRENT FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 636
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 312
LENGTH: 111999
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(111999)
OTHER INFORMATION: n = A,T,C or G
US-60-212-664-312

Alignment Scores:

Pred. No.:	5.55e-15	
Score:	200.00	Length:	111999
Percent Similarity:	97.44%	Matches:	38
		Conservative:	0
Best Local Similarity:	97.44%	Mismatches:	1
Query Match:	85.84%	Indels:	0
DB:		Gaps:	0

US-10-045-180A-5 (1-44) x US-60-212-664-312 (1-111999)

```
Qy     1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 67705 GAGCCCGCTCCAGGCAAGAGCTGATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGAC 67764

Qy    21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerSerLeuGlnValProGly 39
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 67765 CAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTTCTTCAGGTTCCAGGT 67821
```

RESULT 11
US-60-171-481-272/c
; Sequence 272, Application US/60171481
; GENERAL INFORMATION:
APPLICANT: Bonazzi, Vivien

; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS AND
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000169
; CURRENT APPLICATION NUMBER: US/60/171,481
; CURRENT FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 1898
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 272
; LENGTH: 345
; TYPE: DNA
; ORGANISM: Human
US-60-171-481-272

Alignment Scores:

Pred. No.:	4.25e-16	
Score:	187.00	Length:	345
Percent Similarity:	95.00%	Matches:	38
		Conservative:	0
Best Local Similarity:	95.00%	Mismatches:	1
Query Match:	80.26%	Indels:	1
DB:		Gaps:	0

US-10-045-180A-5 (1-44) x US-60-171-481-272 (1-345)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGln-LysGlnProProAlaAspAs 20
       |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db 219 GAGCCCGCTCCAGGCAAGAGCTGATGAGATGCCAAATAAGCAGCCTCCAGCAGATGA 160

Qy  20 pGlnAspValValIleTyrPheSerGlyAspAspSerCysSerSerLeuGlnValProGly 39
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 159 CCAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTTCTTCAGGTTCCAGGT 102
```

RESULT 12
US-60-171-481-777/c
; Sequence 777, Application US/60171481
; GENERAL INFORMATION:
APPLICANT: Bonazzi, Vivien
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: CL000169
CURRENT APPLICATION NUMBER: US/60/171,481
CURRENT FILING DATE: 1999-12-22
NUMBER OF SEQ ID NOS: 1898
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 777
LENGTH: 345
TYPE: DNA
ORGANISM: Human
US-60-171-481-777

Alignment Scores:

Pred. No.:	4.25e-16	
Score:	187.00	Length:	345
Percent Similarity:	95.00%	Matches:	38
		Conservative:	0
Best Local Similarity:	95.00%	Mismatches:	1
Query Match:	80.26%	Indels:	1
DB:		Gaps:	0

US-10-045-180A-5 (1-44) x US-60-171-481-777 (1-345)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGln-LysGlnProProAlaAspAs 20
       |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db 219 GAGCCCGCTCCAGGCAAGAGCTGATGAGATGCCAAATAAGCAGCCTCCAGCAGATGA 160

Qy  20 pGlnAspValValIleTyrPheSerGlyAspAspSerCysSerSerLeuGlnValProGly 39
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 159 CCAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTTCTTCAGGTTCCAGGT 102
```

RESULT 13
US-09-321-214-10452/c
; Sequence 10452, Application US/09321214
; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Ford, John
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-740
; CURRENT APPLICATION NUMBER: US/09/321,214
; CURRENT FILING DATE: 1999-05-26
; EARLIER APPLICATION NUMBER: US 60/088,041
; EARLIER FILING DATE: 1998-06-02
; NUMBER OF SEQ ID NOS: 31906
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10452
; LENGTH: 338
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-321-214-10452

Alignment Scores:
Pred. No.:
Score: 9.74e-07
Percent Similarity: 72.73%
Best Local Similarity: 59.09%
Query Match: 51.50%
DB: 19

Length: 338
Matches: 26
Conservative: 6
Mismatches: 12
Indels: 0
Gaps: 0

US-10-045-180A-5 (1-44) x US-09-321-214-10452 (1-338)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
         |||||||||||||||||:::||||||:::|||||||::
Db 260 GAGCCACTCCAGGCAAATTGCTGATGAGGTTACAGCCCAGGAGCAGCCTGGAGCAGATGAT 201

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
         ||||::||||||||||||||
Db 200 CAGGAAGTGGTTGAATTCCTTTGCATGGGATGAAAAGAGTCCTCTTCAGGTTTCAGGCTCA 141

Qy  41 ThrLysGlyLeu 44
         ::||||||||||
Db 140 AGGAGGGGCTTG 129
```

RESULT 14
US-09-516-335-10452/c
Sequence 10452, Application US/09516335
GENERAL INFORMATION:
; APPLICANT: Arterburn, Matthew
; APPLICANT: Asghari, Vida
; APPLICANT: Damavandi, Simin
; APPLICANT: Dickson, Mark
; APPLICANT: Drake, Jim
; APPLICANT: Drmanac, Radoje
; APPLICANT: Engleman, Carrie
; APPLICANT: Faulkner, Brandy
; APPLICANT: Fox, Melvin
; APPLICANT: Garcia, Veronica
; APPLICANT: Giedt, Gretchen
; APPLICANT: Jessen, Aaron
; APPLICANT: Jomek, Leni
; APPLICANT: Jones, Lee
; APPLICANT: Kita, David
; APPLICANT: Labat, Ivan
; APPLICANT: Laroya, Mimi
; APPLICANT: Lomelli, Michelle
; APPLICANT: Nelken, Sarah
; APPLICANT: Nguyen, Kody
; APPLICANT: Nguyen, Lynne
; APPLICANT: Nguyen, Phuong
; APPLICANT: Nogra, Margie

; APPLICANT: Palencia, Servando
; APPLICANT: Raisi, Fariba
; APPLICANT: Randhwa, Gurpreet
; APPLICANT: Sidhu, Navjiwan
; APPLICANT: Smith, Benjamin
; APPLICANT: Smythe, Ashleigh
; APPLICANT: Tkach, Joe
; APPLICANT: Tran, Lien
; APPLICANT: Verna, Ron
; APPLICANT: Wachter, Adam
; APPLICANT: Wu, James
; APPLICANT: Yim, Kenneth
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 740CIP
; CURRENT APPLICATION NUMBER: US/09/516,335
; CURRENT FILING DATE: 2000-03-01
; EARLIER APPLICATION NUMBER: 09/321,214
; EARLIER FILING DATE: 1999-05-26
; EARLIER APPLICATION NUMBER: 60/088,041
; EARLIER FILING DATE: 1998-06-02
; NUMBER OF SEQ ID NOS: 31906
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10452
; LENGTH: 338
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-516-335-10452

Alignment Scores:
Pred. No.:
Score: 9.74e-07
Percent Similarity: 72.73%
Best Local Similarity: 59.09%
Query Match: 51.50%
DB: 22

Length: 338
Matches: 26
Conservative: 6
Mismatches: 12
Indels: 0
Gaps: 0

US-10-045-180A-5 (1-44) x US-09-516-335-10452 (1-338)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
         |||||||||||||||||:::||||||:::|||||||::
Db 260 GAGCCACTCCAGGCAAATTGCTGATGAGGTTACAGCCCAGGAGCAGCCTGGAGCAGATGAT 201

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
         ||||::||||||||||||||
Db 200 CAGGAAGTGGTTGAATTCCTTTGCATGGGATGAAAAGAGTCCTCTTCAGGTTTCAGGCTCA 141

Qy  41 ThrLysGlyLeu 44
         ::||||||||||
Db 140 AGGAGGGGCTTG 129
```

RESULT 15
US-09-733-811-10452/c
Sequence 10452, Application US/09733811
GENERAL INFORMATION:
; APPLICANT: Arterburn, Matthew
; APPLICANT: Asghari, Vida
; APPLICANT: Damavandi, Simin
; APPLICANT: Dickson, Mark
; APPLICANT: Drake, Jim
; APPLICANT: Drmanac, Radoje
; APPLICANT: Engleman, Carrie
; APPLICANT: Faulkner, Brandy
; APPLICANT: Fox, Melvin
; APPLICANT: Garcia, Veronica
; APPLICANT: Giedt, Gretchen
; APPLICANT: Jessen, Aaron
; APPLICANT: Jomek, Leni
; APPLICANT: Jones, Lee
; APPLICANT: Kita, David
; APPLICANT: Labat, Ivan
; APPLICANT: Laroya, Mimi
; APPLICANT: Lomelli, Michelle

```
;  APPLICANT:  Nelken, Sarah
;  APPLICANT:  Nguyen, Kody
;  APPLICANT:  Nguyen, Lynne
;  APPLICANT:  Nguyen, Phuong
;  APPLICANT:  Nogra, Margie
;  APPLICANT:  Palencia, Servando
;  APPLICANT:  Raisi, Fariba
;  APPLICANT:  Randhwa, Gurpreet
;  APPLICANT:  Sidhu, Navjiwan
;  APPLICANT:  Smith, Benjamin
;  APPLICANT:  Smythe, Ashleigh
;  APPLICANT:  Tkach, Joe
;  TITLE OF INVENTION:  NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE:  740CIP
;  CURRENT APPLICATION NUMBER: US/09/733,811
;  CURRENT FILING DATE:  2000-12-08
;  PRIOR APPLICATION NUMBER: 09/321,214
;  PRIOR FILING DATE:  1999-05-26
;  PRIOR APPLICATION NUMBER:  60/088,041
;  PRIOR FILING DATE:  1998-06-02
;  NUMBER OF SEQ ID NOS:  31906
;  SOFTWARE:  FastSEQ for Windows Version 3.0
;  SEQ ID NO 10452
;    LENGTH:  338
;    TYPE:  DNA
;    ORGANISM:  Homo sapiens
US-09-733-811-10452

Alignment Scores:
Pred. No.:
Score:                          9.74e-07      Length:        338
Percent Similarity:             120.00        Matches:       26
Best Local Similarity:          72.73%        Conservative:  6
Query Match:                    59.09%        Mismatches:    12
DB:                             51.50%        Indels:        0
                                32            Gaps:          0

US-10-045-180A-5 (1-44) x US-09-733-811-10452 (1-338)

Qy       1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp    20
           |||||||||||||||||| |||||  ||||||:::  ||||||||||||:||||||||||
Db     260 GAGCCACTCCAGGCAATTGCTGATGAGGTTACAGCCCAGGAGCAGCCCTGGAGCAGAGTGAT   201

Qy      21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer   40
           |||||:::|||||||||  ||||||||||  :::     ||||||||:|||||||||
Db     200 CAGGAAGTGGTTGATTCCTTTGCATGGGATGGATGAAAGAGCTCCTCTTCAGGTTTCAGGCTCA   141

Qy      41 ThrLysGlyLeu    44
           :::||||||||
Db     140 AGGAGGGGGCTTG    129

Search completed: December 17, 2003, 21:11:01
Job time : 2320.08 secs
```

OM protein - nucleic search, using frame_plus_p2n model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.6

Run on: December 17, 2003, 14:50:33 ; Search time 44 Seconds
441.383 Million cell updates/sec
(without alignments)

Sequence: 1 EPLQARAHEMPAQKQPPADD.........IYFSGDDSCSLQVPGSTKGL 44

Perfect score: 233

Title: US-10-045-180A-5

Scoring table: BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop   6.0 ; Fgapext 7.0
Delop   6.0 ; Delext   7.0

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145024_12553/app_query.fasta_1.860
-DB=Issued_Patents_NA:*
-QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCI=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10045180 @CGN 1 1_193 @runat_17122003_145024_12553 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7
-FGAPEXT=7 -YGAPEXT=0.5 -YGAPOP=10 -FGAPOP=6

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match Length	DB	ID	Description
1	233	100.0	453	4	US-09-486-580A-2	Sequence 2, Appli
2	209	89.7	4415	4	US-09-486-580A-1	Sequence 1, Appli
3	101	43.3	452	1	US-08-158-189-6	Sequence 6, Appli
4	101	43.3	452	1	US-09-016-434-1198	Sequence 1198, Ap
5	86	36.9	243	4	US-09-309-487-28	Sequence 28, Appl
6	86	36.9	243	4	US-09-967-808-28	Sequence 28, Appl
7	83	35.6	495	4	US-09-309-487-15	Sequence 15, Appl
8	83	35.6	495	4	US-09-967-808-15	Sequence 15, Appl
9	82	35.2	500	4	US-09-309-487-13	Sequence 13, Appl
10	82	35.2	500	4	US-09-967-808-13	Sequence 13, Appl
11	80	34.3	424	1	US-08-158-189-4	Sequence 4, Appl
12	80	34.3	2553	4	US-09-309-487-25	Sequence 25, Appl
13	80	34.3	2553	4	US-09-967-808-25	Sequence 25, Appl
14	76	32.6	2525	4	US-09-309-487-24	Sequence 24, Appl
15	76	32.6	2525	4	US-09-967-808-24	Sequence 24, Appl
16	76	32.6	2880	1	US-08-158-189-1	Sequence 1, Appli
17	74	31.8	445	1	US-08-486-013-66	Sequence 66, Appl
18	74	31.8	445	2	US-08-482-279-66	Sequence 66, Appl
19	74	31.8	445	2	US-08-342-268-66	Sequence 66, Appl
20	74	31.8	445	1	US-09-397-386-66	Sequence 66, Appl
21	74	31.8	445	4	US-09-015-968-66	Sequence 66, Appl
22	74	31.8	2408	1	US-08-486-013-69	Sequence 69, Appl
23	74	31.8	2408	2	US-08-482-279-69	Sequence 69, Appl
24	74	31.8	2408	2	US-08-342-268-69	Sequence 69, Appl
25	74	31.8	2408	3	US-09-015-968-69	Sequence 69, Appl
26	74	31.8	2408	4	US-09-397-386-69	Sequence 69, Appl
27	72.5	31.1	218	1	US-08-158-189-38	Sequence 38, Appl
28	72	30.9	2457	1	US-08-486-013-68	Sequence 68, Appl
29	72	30.9	2457	2	US-08-482-279-68	Sequence 68, Appl
30	72	30.9	2457	2	US-08-342-268-68	Sequence 68, Appl
31	72	30.9	2457	3	US-09-015-968-68	Sequence 68, Appl
32	72	30.9	2457	4	US-09-397-386-68	Sequence 68, Appl
33	72	30.9	2551	2	US-08-486-013-70	Sequence 70, Appl
34	72	30.9	2551	2	US-08-482-279-70	Sequence 70, Appl
35	72	30.9	2551	2	US-08-342-268-70	Sequence 70, Appl
36	72	30.9	2551	2	US-09-015-968-70	Sequence 70, Appl
37	72	30.9	2551	4	US-09-397-386-70	Sequence 70, Appl
38	71	30.5	445	1	US-08-486-013-65	Sequence 65, Appl
39	71	30.5	445	1	US-08-486-013-67	Sequence 67, Appl
40	71	30.5	445	2	US-09-397-386-67	Sequence 67, Appl
41	71	30.5	445	2	US-08-482-279-65	Sequence 65, Appl
42	71	30.5	445	2	US-08-482-279-67	Sequence 67, Appl
43	71	30.5	445	2	US-08-342-268-65	Sequence 65, Appl
44	71	30.5	445	2	US-08-342-268-67	Sequence 67, Appl
45	71	30.5	445	3	US-09-015-968-65	Sequence 65, Appl

ALIGNMENTS

RESULT 1
US-09-486-580A-2
Sequence 2, Application US/09465080A
Patent No. 6329340
GENERAL INFORMATION:
APPLICANT: GENSET SA
TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 550 West C Street
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/486,580A
FILING DATE: FEBRUARY 25, 2000
ATTORNEY/AGENT INFORMATION:
NAME: Hart, Daniel
REGISTRATION NUMBER: 40,637
REFERRENCE/DOCKET NUMBER: GENSET.064C1
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 453 BASE PAIRS
TYPE: NUCLEOTIDE
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA

; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-486-580A-2

Alignment Scores:
Pred. No.:
Score: 233.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 4

US-10-045-180A-5 (1-44) x US-09-486-580A-2 (1-453)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnIleLysGlnProProAlaAspAsp 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109 GAGCCCGCTCCAGGCAAGAGCTTCATGAGGATGCCAGCCCAGAGAAGCAGCCTCCAGCAGATGAC 168

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 169 CAGGATGGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGCTCA 228

Qy   41 ThrLysGlyLeu 44
         ||||||||||||
Db 229 ACAAAGGGCTTG 240

RESULT 2
US-09-486-580A-1
; Sequence 1, Application US/0948580A
; Patent No. 6329340
; GENERAL INFORMATION:
; APPLICANT: GENSET SA
; TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAc
; TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
; TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/486,580A
;   FILING DATE: FEBRUARY 25, 2000
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 4415 BASE PAIRS
;   TYPE: NUCLEOTIDE
;   STRANDEDNESS: DOUBLE
;   TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
; FEATURE:
;   NAME/KEY: Exon 1
;   LOCATION: 1836..1874
; FEATURE:
;   NAME/KEY: Exon 2
;   LOCATION: 3394..3577
; FEATURE:
;   NAME/KEY: Exon 3
;   LOCATION: 4161..4380

; FEATURE:
;   NAME/KEY: start CDS
;   LOCATION: 3406..3408
; FEATURE:
;   NAME/KEY: stop CDS
;   LOCATION: 4276..4278
; FEATURE:
;   NAME/KEY: polyadenylation site
;   LOCATION: 4374..4379
US-09-486-580A-1

Alignment Scores:
Pred. No.:
Score: 209.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 4

US-10-045-180A-5 (1-44) x US-09-486-580A-1 (1-4415)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnIleLysGlnProProAlaAspAsp 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3463 GAGCCCGCTCCAGGCAAGAGCTTCATGAGGATGCCAGCCCAGAGAAGCAGCCTCCAGCAGATGAC 3522

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGly 39
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3523 CAGGATGGTGGTCATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGT 3579

RESULT 3
US-08-158-189-6
; Sequence 6, Application US/0815819
; Patent No. 5641497
; GENERAL INFORMATION:
; APPLICANT: Bevins, Charles L.
; APPLICANT: Jones, Douglas E.
; TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
;   STREET: One Liberty Place - 46th Floor
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: U.S.A.
;   ZIP: 19103
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/158,189
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/07/888,232
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Johnson, Philip S.
;   REGISTRATION NUMBER: 27,200
;   REFERENCE/DOCKET NUMBER: CH-0219
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-568-3100
;   TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 452 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
; FEATURE:
;   NAME/KEY: CDS

; LOCATION: 19..321
US-08-158-189-6

Alignment Scores:
Pred. No.:
Score: 2.38e-06
Percent Similarity: 101.00
Best Local Similarity: 71.43%
Query Match: 50.00%
DB: 43.35%
                                    1

Length: 452
Matches: 21
Conservative: 9
Mismatches: 12
Indels: 0
Gaps: 0

US-10-045-180A-5 (1-44) x US-08-158-189-6 (1-452)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProAlaAspAsp  20
         :::||||||||::||||||||||||:::|||:::||:::|||
Db  97 GATCCACTGCAGGCAAAAAGCTTATGAGGCTGAGTCCCCAGGAGCAGCGGCAAATGAC 156

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer  40
         ||||||:::  |||:::    |||:::
Db 157 CAGGACTTTGCCGTCTCCTTTGCAGAGGATGCAAGCTCAAGTCTTAGAGCTTTGGGCTCA 216

Qy  41 ThrLys  42
         |||:::
Db 217 ACAAGG 222
```

RESULT 4
US-09-016-434-1198
; Sequence 1198, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Incyte Pharmaceuticals, Inc.
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1198:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 452 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 9181546
US-09-016-434-1198

Alignment Scores:
Pred. No.:
Score: 2.38e-06
Percent Similarity: 101.00
Best Local Similarity: 71.43%
Query Match: 50.00%
DB: 43.35%
                                    1

Length: 452
Matches: 21
Conservative: 9
Mismatches: 12
Indels: 0
Gaps: 0

US-10-045-180A-5 (1-44) x US-09-016-434-1198 (1-452)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProAlaAspAsp  20
         :::||||||||::||||||||||||:::|||:::||:::|||
Db  97 GATCCACTGCAGGCAAAAAGCTTATGAGGCTGAGTCCCCAGGAGCAGCGGCAAATGAC 156

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer  40
         ||||||:::  |||:::    |||:::
Db 157 CAGGACTTTGCCGTCTCCTTTGCAGAGGATGCAAGCTCAAGTCTTAGAGCTTTGGGCTCA 216

Qy  41 ThrLys  42
         |||:::
Db 217 ACAAGG 222
```

RESULT 5
US-09-309-487-28
; Sequence 28, Application US/09309487
; Patent No. 6335318
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Tang, Yi-Quan
; APPLICANT: Yuan, Jun
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antimicrobial Theta Defensins and Methods of Using Same
; FILE REFERENCE: P-UC 3095
; CURRENT APPLICATION NUMBER: US/09/309,487
; CURRENT FILING DATE: 1999-05-10
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 28
; LENGTH: 243
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(237)
US-09-309-487-28

Alignment Scores:
Pred. No.:
Score: 0.000184
Percent Similarity: 86.00
Best Local Similarity: 61.36%
Query Match: 45.45%
DB: 36.91%
                                    4

Length: 243
Matches: 20
Conservative: 7
Mismatches: 17
Indels: 0
Gaps: 0

US-10-045-180A-5 (1-44) x US-09-309-487-28 (1-243)

```
Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProAlaAspAsp  20
         |||||||||||||||:::|||||||:::||||||||||||||
Db  67 GAGGCACGTCAGGCAAGAGCTGATGAAGCTGCCCCCAGCAGCAGCAGATGAT 126

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer  40
         |||:::  |||:::    :::
Db 127 CAGGGAAATGGCTCAATCCTTTACATGGCCTGAAAACGGCGCGTCTTCCACTTTCAGAGTCA 186

Qy  41 ThrLysGlyLeu  44
         ||||||||||||
Db 187 GCGAAAGGCTTG 198
```

RESULT 6
US-09-967-808-28
; Sequence 28, Application US/09967808
; Patent No. 6514727
; GENERAL INFORMATION:

APPLICANT: Selsted, Michael E.
APPLICANT: Tang, Yi-Quan
APPLICANT: Yuan, Jun
APPLICANT: Ouellette, Andre J.
TITLE OF INVENTION: Antimicrobial Theta Defensins and Methods of Using
FILE REFERENCE: P-UC 3095
CURRENT APPLICATION NUMBER: US/09/967,808
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US/09/309,487
PRIOR FILING DATE: 1999-05-10
NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 28
LENGTH: 243
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (7)..(237)

US-09-967-808-28

Alignment Scores:
Pred. No.:
Score: 0.000184
Percent Similarity: 86.00
Best Local Similarity: 61.36%
Query Match: 45.45%
DB: 36.91%
                                          4

US-10-045-180A-5 (1-44) x US-09-967-808-28 (1-243)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
        ||||||||||||||||||||:::|||||||||||||||||:::|||||||||
Db   67 GAGGCACGTCAGGCAAGAGCTGATGAAGCTGCCGCCCAGCAGCAGCCTGGAGCAGATGAT 126

Qy   21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
        |||:::||||||||||||||||
Db 127 CAGGGGAATGGCCTCATTCCTTTACACGGCCTGAAAAACGCCCGGCTCTTCCACTTTCAGAGTCA 186

Qy   41 ThrLysGlyLeu 44
        |||||||
Db 187 GCGGAAAGGCTTG 198

RESULT 7
US-09-309-487-15
Sequence 15, Application US/09309487
Patent No. 6335318
GENERAL INFORMATION:
CURRENT FILING DATE: 1999-05-10
NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 495
APPLICANT: Tang, Yi-Quan
APPLICANT: Ouellette, Andre J.
ORGANISM: Macaca mulatta
FEATURE:
NAME/KEY: CDS
LOCATION: (90)..(320)

US-09-309-487-15

```
; APPLICANT: Yuan, Jun
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antimicrobial Theta Defensins and Methods of Using Same
; FILE REFERENCE: P-UC 3095
; CURRENT APPLICATION NUMBER: US/09/309,487
; CURRENT FILING DATE: 1999-05-10
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 13
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Macaca mulatta
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (95)..(325)
US-09-309-487-13

Alignment Scores:
Pred. No.:                0.0178
Score:                    82.00
Percent Similarity:       59.09%
Best Local Similarity:    43.18%
Query Match:              35.19%

                          Length:        500
                          Matches:        19
                          Conservative:    7
                          Mismatches:     18
                          Indels:          0
                          Gaps:            0

US-10-045-180A-5 (1-44) x US-09-309-487-13 (1-500)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
         |||    |:|:|| ||||| ::::|||||::|:||||||
Db 155 GAGGCACGTCAGGCAAGAGCTGATGAAGCTGCCCGCCCAGCAGCAGCCTGAACAGATGAT 214

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
         ||||:::      ::|||    |
Db 215 CAGGGAAATGGCTCATTCCTTTACATGGGCCTGAAAACGCCGCGCTCTTCCACTTTCAGAGTCA 274

Qy  41 ThrLysGlyLeu 44
         |||||||||
Db 275 GCGAAAGGCTTG 286
```

RESULT 10
US-09-967-808-13

```
; Sequence 13, Application US/09967808
; Patent No. 6514727
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Tang, Yi-Quan
; APPLICANT: Yuan, Jun
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antimicrobial Theta Defensins and Methods of Using
; TITLE OF INVENTION: Same
; FILE REFERENCE: P-UC 3095
; CURRENT APPLICATION NUMBER: US/09/967,808
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US/09/309,487
; PRIOR FILING DATE: 1999-05-10
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 13
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Macaca mulatta
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (95)..(325)
US-09-967-808-13

Alignment Scores:
Pred. No.:                0.0178
Score:                    82.00
Percent Similarity:       59.09%
Best Local Similarity:    43.18%
Query Match:              35.19%

                          Length:        500
                          Matches:        19
                          Conservative:    7
                          Mismatches:     18
                          Indels:          0
                          Gaps:            0

US-10-045-180A-5 (1-44) x US-09-967-808-13 (1-500)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
         |||    |:|:|| ||||| ::::|||||::|:||||||
Db 155 GAGGCACGTCAGGCAAGAGCTGATGAAGCTGCCCGCCCAGCAGCAGCCTGAACAGATGAT 214

Qy  21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
         ||||:::      ::|||    |
Db 215 CAGGGAAATGGCTCATTCCTTTACATGGGCCTGAAAACGCCGCGCTCTTCCACTTTCAGAGTCA 274

Qy  41 ThrLysGlyLeu 44
         |||||||||
Db 275 GCGAAAGGCTTG 286
```

RESULT 11
US-08-158-189-4

```
; Sequence 4, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
; APPLICANT: Bevins, Charles L.
; APPLICANT: Jones, Douglas E.
; TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 564149?ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,189
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,232
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Johnson, Philip S.
; REGISTRATION NUMBER: 27,200
; REFERENCE/DOCKET NUMBER: CH-0219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 424 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10..294
US-08-158-189-4

Alignment Scores:
Pred. No.:                0.00287
Score:                    80.00
Percent Similarity:       60.00%
Best Local Similarity:    45.00%
Query Match:              34.33%

                          Length:        424
                          Matches:        18
                          Conservative:    6
                          Mismatches:     16
                          Indels:          0
                          Gaps:            0

US-10-045-180A-5 (1-44) x US-08-158-189-4 (1-424)

Qy   1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
```

Db
Qy 67 GAGTCACTCCAGGAAAGAGAGAGTCTGATGAGGCTACAACCCAGAAGCAGTCTGGGGGAAGACAAC
Qy 21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
Db 127 CAGGACCTTGCTATCTCCTTTGCAGGAAAATGGACTCTCTGCTCTTAGAACCTCAGGTTCT 186

RESULT 12
US-09-309-487-25
; Sequence 25, Application US/09309487
; Patent No. 6335318
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Tang, Yi-Quan
; APPLICANT: Yuan, Jun
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antimicrobial Theta Defensins and Methods of Using Same
; FILE REFERENCE: P-UC 3095
; CURRENT APPLICATION NUMBER: US/09/309,487
; CURRENT FILING DATE: 1999-05-10
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 25
; LENGTH: 2553
; TYPE: DNA
; ORGANISM: Macaca mulatta
US-09-309-487-25

Alignment Scores:
Pred. No.:   Length: 2553
Score: 0.027   Matches: 18
Percent Similarity: 80.00   Conservative: 7
Best Local Similarity: 60.98%   Mismatches: 16
Query Match: 43.90%   Indels: 0
DB: 34.33%   Gaps: 0
                        4

US-10-045-180A-5 (1-44) x US-09-309-487-25 (1-2553)

Qy 1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
Db 1521 GAGGCACGTCAGGCAGCAAGAGCTGATGAAGCTCCGCGCCCCAGCAGCAGCCTGGAGCAGAGATGAT 1580

Qy 21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
Db 1581 CAGGGAAATGGCTCATTCCTTTACACGGCGCCTGAAAAACGCGCGTCTTCCGGCTTTCAGGTGAG 1640

Qy 41 Thr 41
Db 1641 ACA 1643

RESULT 13
US-09-967-808-25
; Sequence 25, Application US/0996708
; Patent No. 6514727
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Tang, Yi-Quan
; APPLICANT: Yuan, Jun
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antimicrobial Theta Defensins and Methods of Using
; TITLE OF INVENTION: Same
; FILE REFERENCE: P-UC 3095
; CURRENT APPLICATION NUMBER: US/09/967,808
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US/09/309,487
; PRIOR FILING DATE: 1999-05-10
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 25
; LENGTH: 2553
; TYPE: DNA
; ORGANISM: Macaca mulatta
US-09-967-808-25

US-10-045-180A-5 (1-44) x US-09-967-808-25 (1-2553)

Qy 1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
Db 1521 GAGGCACGTCAGGCAGCAAGAGCTGATGAAGCTCCGCGCCCCAGCAGCAGCCTGGAGCAGAGATGAT 1580

Qy 21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
Db 1581 CAGGGAAATGGCTCATTCCTTTACACGGCGCCTGAAAAACGCGCGTCTTCCGGCTTTCAGGTGAG 1640

Qy 41 Thr 41
Db 1641 ACA 1643

RESULT 14
US-09-309-487-24
; Sequence 24, Application US/09309487
; Patent No. 6335318
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Tang, Yi-Quan
; APPLICANT: Yuan, Jun
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antimicrobial Theta Defensins and Methods of Using Same
; FILE REFERENCE: P-UC 3095
; CURRENT APPLICATION NUMBER: US/09/309,487
; CURRENT FILING DATE: 1999-05-10
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 24
; LENGTH: 2525
; TYPE: DNA
; ORGANISM: Macaca mulatta
US-09-309-487-24

Alignment Scores:
Pred. No.:   Length: 2525
Score: 0.104   Matches: 17
Percent Similarity: 76.00   Conservative: 7
Best Local Similarity: 58.54%   Mismatches: 17
Query Match: 41.46%   Indels: 0
DB: 32.62%   Gaps: 0
                        4

US-10-045-180A-5 (1-44) x US-09-309-487-24 (1-2525)

Qy 1 GluProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp 20
Db 1498 GAGGCACGTCAGGCAGCAAGAGCTGATGAAGCTGCGCGCCCCAGCAGCCTGGAACAGAGAT 1557

Qy 21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer 40
Db 1558 CAGGGAAATGGCTCATTCCTTTACATGGCCTGAAAAACGCGCGTCTTCCACTTTCAGGTGAG 1617

Qy 41 Thr 41
Db 1618 ACA 1620

RESULT 15
US-09-967-808-24
; Sequence 24, Application US/0996708
; Patent No. 6514727
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Tang, Yi-Quan

```
;   APPLICANT:  Yuan, Jun
;   APPLICANT:  Ouellette, Andre J.
;   TITLE OF INVENTION: Antimicrobial Theta Defensins and Methods of Using
;   TITLE OF INVENTION:  Same
;   FILE REFERENCE: P-UC 3095
;   CURRENT APPLICATION NUMBER: US/09/967,808
;   CURRENT FILING DATE:  2001-09-26
;   PRIOR APPLICATION NUMBER: US/09/309,487
;   PRIOR FILING DATE: 1999-05-10
;   NUMBER OF SEQ ID NOS: 31
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 24
;     LENGTH: 2525
;     TYPE: DNA
;     ORGANISM: Macaca mulatta
;   US-09-967-808-24

Alignment Scores:
Pred. No.:
Score:               0.104           Length:        2525
Percent Similarity:  76.00           Matches:       17
Best Local Similarity: 58.54%        Conservative:  7
Query Match:         41.46%          Mismatches:    17
                     32.62%          Indels:        0
                     4               Gaps:          0
DB:

US-10-045-180A-5  (1-44) x US-09-967-808-24  (1-2525)

Qy        1 GlyProLeuGlnAlaArgAlaHisGluMetProAlaGlnLysGlnProProAlaAspAsp  20
            |||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db     1498 GAGGCACGTCAGGCAAGAGCTGATGAAGCTGCCCGCCAGCAGCCTGGAACAGATGAT 1557

Qy       21 GlnAspValValIleTyrPheSerGlyAspAspSerCysSerLeuGlnValProGlySer  40
            |||:||                   ::::|||      ::::|||        |||
Db     1558 CAGGGAATGGCTCATTCCTTTACATGGCCTGAAAACGCCGCTCTTCCACTTTCAGGTGAG 1617

Qy       41 Thr 41
            |||
Db     1618 ACA 1620
```

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: December 17, 2003, 14:50:33 ; Search time 886.963 seconds
(without alignments)
1429.821 Million cell updates/sec

Title:
Perfect score: US-10-045-180a-6
Sequence: 191
Sequence: 1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31

Scoring table: BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop 6.0 ; Delext 7.0

Searched: 2888711 seqs, 2045813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145023_12528/app_query.fasta_1.860
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10045180 @CGN_1_5283 @runat_17122003_145023_12528 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*

```
 1:   gb_ba:*
 2:   gb_bl:*
 3:   gb_in:*
 4:   gb_Om:*
 5:   gb_ov:*
 6:   gb_pat:*
 7:   gb_ph:*
 8:   gb_pl:*
 9:   gb_pr:*
10:   gb_ro:*
11:   gb_sts:*
12:   gb_sy:*
13:   gb_un:*
14:   gb_vi:*
15:   gb_ba:*
16:   em_fun:*
17:   em_hum:*
18:   em_in:*
19:   em_mu:*
20:   em_om:*
21:   em_or:*
22:   em_ov:*
23:   em_pat:*
24:   em_ph:*
25:   em_pl:*
26:   em_ro:*
27:   em_sts:*
28:   em_un:*
29:   em_vi:*
30:   em_htg_hum:*
31:   em_htg_inv:*
32:   em_htg_other:*
33:   em_htg_mus:*
34:   em_htg_pln:*
35:   em_htg_rod:*
36:   em_htg_vrt:*
37:   em_htg_mam:*
38:   em_sy:*
39:   em_htgo_hum:*
40:   em_htgo_mus:*
41:   em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB ID	Description
	1	191	100.0	453	6	A98571	A98571 Sequence 2
	2	191	100.0	453	6	BD074746	BD074746 Human def
	3	191	100.0	4415	6	A98570	A98570 Sequence 1
	4	191	100.0	4415	6	BD074745	BD074745 Human def
c	5	191	100.0	128544	9	AF238378	AF238378 Homo sapi
	6	159	83.2	163093	2	AC116559	AC116559 Papio anu
c	7	159	83.2	188604	2	AC116558	AC116558 Papio anu
c	8	101	52.9	271979	2	AC098442	AC098442 Rattus no
c	9	101	52.9	301130	2	AC128185	AC128185 Rattus no
	10	89	46.6	347	9	AF188268	AF188268 Macaca mu
	11	89	46.6	435	9	AF184160	AF184160 Macaca mu
	12	87	45.5	433	9	AF184159	AF184159 Macaca mu
85.5	13	85.5	44.8	1260	10	RNU50356	U50356 Rattus norv
	14	84	44.0	494	10	RNU16684	U16684 Rattus norv
	15	84	44.0	504	10	RNU16683	U16683 Rattus norv
c	16	84	44.0	3003	10	RRU50354	U50354 Rattus norv
	17	84	44.0	3046	10	RNU50353	U50353 Rattus norv
	18	84	44.0	3400	10	RNU50355	U50355 Rattus norv
c	19	84	44.0	242829	2	AC114391	AC114391 Rattus no
c	20	82	42.9	67979	2	AC113099	AC113099 Mus muscu
c	21	81	42.4	347	9	AF188270	AF188270 Macaca mu
	22	81	42.4	485	4	RABDNF4A	RABDNF4A Oryctolagu
	23	81	42.4	90	6	M64601	M64601 Oryctolagus
	24	80	41.9	108	6	BD128611	BD128611 Method fo
	25	80	41.9	108	6	I32861	I32861 Sequence 43
c	26	80	41.9	108	6	I32862	I32862 Sequence 44
c	27	80	41.9	109	6	I32878	I32878 Sequence 61
c	28	80	41.9	316	9	HUMDEF3A	HUMDEF3A Homo sapien
	29	80	41.9	448	9	HUMDEF1A	M21131 Homo sapien
	30	80	41.9	451	9	HUMDEFB	M21130 Human neutr
	31	80	41.9	464	9	HSDEF3	M23281 Human defen
	32	80	41.9	478	6	BD028213	X13621 Human HNP-3
	33	80	41.9	482	9	HSHP1	BD028213 Sequence
	34	80	41.9	498	9	BD174049	X52053 Human mRNA
	35	80	41.9	498	9	HUMDEF1AA	BD174049 Method of
	36	80	41.9	509	6	BD023843	M26602 Human defen
	37	80	41.9	514	6	AX405718	BD023843 Sequence
	38	80	41.9	543	6	AX588765	AX405718 Sequence
	39	80	41.9	556	6	BC027917	AX588765 Sequence
c	40	80	41.9	1097	6	I30337	BC027917 Homo sapi
c	41	80	41.9	1097	6	I96061	I30337 Sequence 5
c	42	80	41.9	2147	17	AF119902	I96061 Sequence 5
	43	80	41.9	3710	6	BD174050	AF119902 Homo sapi
	44	80	41.9	3710	6	HUMNTRI	BD174050 Method of
	45	80	41.9	3710	9	HUMNTRIII	L12690 Human neutr
							L12691 Human neutr

ALIGNMENTS

RESULT 1

LOCUS   A98571   453 bp   DNA   linear   PAT 26-JAN-2000
DEFINITION   Sequence 2 from Patent WO9911663.
ACCESSION   A98571
VERSION   A98571.1 GI:6781627
KEYWORDS   .
SOURCE   unidentified
  ORGANISM   unidentified
             unclassified.
REFERENCE   1 (bases 1 to 453)
  AUTHORS   Bougueleret,L. and Chumakov,I.
  TITLE   HUMAN DEFENSIN DEF-X, GENE AND DNAC, COMPOSITION CONTAINING SAME
             AND DIAGNOSTIC AND THERAPEUTIC APPLICATIONS
  JOURNAL   Patent: WO 9911663-A 2 11-MAR-1999;
             BOUGUELERET LYDIE (FR); CHUMAKOV ILYA (FR)
FEATURES   Location/Qualifiers
     source   1..453
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT   107 a   121 c   97 g   128 t
ORIGIN

US-10-045-180A-6 (1-31) x A98571 (1-453)

Alignment Scores:
Pred. No.:   5.7e-19
Score: 191.00   Length: 453
Percent Similarity: 100.00%   Matches: 31
Percent Identity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 6   Gaps: 0

Qy   1 IleLeuGlyGlyLeuArgGlyTyrProIleCysTyr 31
         ||||||||||||||||||||||||||||||||||||
Db   241 ATCTGCCATTGCAGAGTACTATACTGCATTTTCGGAGAACATCTTGGTGGGGACCTGCTTC 300

Qy   21 IleLeuGlyGlyLeuArgGlyTyrProIleCysTyr 31
         ||||||||||||||||||||||||||||||||
Db   301 ATCCTTGGTGAACCCTACCCAAATCTGCTGCTAC 333

RESULT 2
BD074746
LOCUS   BD074746   453 bp   DNA   linear   PAT 27-AUG-2002
DEFINITION   Human defensin polypeptide Def-X, genome DNA and cDNA, composition
             containing the same, and application to diagnosis and remedy
             treatment.
ACCESSION   BD074746
VERSION   BD074746.1 GI:22620349
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 453)
  AUTHORS   Bougueleret,L. and Shmacov,I.
  TITLE   Human defensin polypeptide Def-X, genome DNA and cDNA, composition
             containing the same, and application to diagnosis and remedy
  JOURNAL   Patent: JP 2001514264-A 2 11-SEP-2001;
             GENSET
COMMENT   OS   Homo sapiens (human)
             PN   JP 2001514264-A/2
             PD   11-SEP-2001
             PF   28-AUG-1998 JP 2000508701
             PR   29-AUG-1997 FR 97/10823
             PI   LYDIE BOUGUELERET, ILYA SHMACOV
             PC   C07K14/435,A01N43/50,A01N63/00,A61K7/00,A61K38/00,A61P29/00,
             PC   A61P35/00,
             PC   A61P37/02,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N15/09, PC
             PC   C12P21/02,
             PC   C12P21/08,C12Q1/68,G01N33/53,A61K37/02,C12N15/00 CC
             CC   Strandedness: Double;
             CC   Topology: Linear;

CC   Human defensin polypeptide Def-X, genome DNA and cDNA, CC
CC   composition
CC   containing the same, and application to diagnosis and remedy
CC   treatment
FH   Key   Location/Qualifiers
FT   source   1..453
                     /organism='Homo sapiens (human)'.
FT   source   1..453
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT   107 a   121 c   97 g   128 t
ORIGIN

US-10-045-180A-6 (1-31) x BD074746 (1-453)

Alignment Scores:
Pred. No.:   5.7e-19
Score: 191.00   Length: 453
Percent Similarity: 100.00%   Matches: 31
Percent Identity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 6   Gaps: 0

Qy   1 IleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyValGlyThrCysPhe 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ATCTGCCATTGCAGAGTACTATACTGCATTTTCGGAGAACATCTTGGTGGGGACCTGCTTC 300

Qy   21 IleLeuGlyGlyLeuArgGlyTyrProIleCysTyr 31
         ||||||||||||||||||||||||||||||||
Db   301 ATCCTTGGTGAACCCTACCCAAATCTGCTGCTAC 333

RESULT 3
A98570
LOCUS   A98570   4415 bp   DNA   linear   PAT 26-JAN-2000
DEFINITION   Sequence 1 from Patent WO9911663.
ACCESSION   A98570
VERSION   A98570.1 GI:6781626
KEYWORDS   .
SOURCE   unidentified
  ORGANISM   unidentified
             unclassified.
REFERENCE   1 (bases 1 to 4415)
  AUTHORS   Bougueleret,L. and Chumakov,I.
  TITLE   HUMAN DEFENSIN DEF-X, GENE AND DNAC, COMPOSITION CONTAINING SAME
             AND DIAGNOSTIC AND THERAPEUTIC APPLICATIONS
  JOURNAL   Patent: WO 9911663-A 1 11-MAR-1999;
             BOUGUELERET LYDIE (FR); CHUMAKOV ILYA (FR)
FEATURES   Location/Qualifiers
     source   1..4415
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
     exon   1836..1874
     exon   3394..3577
     exon   4161..4380
     polyA_site   4374..4379
BASE COUNT   1128 a   1109 c   912 g   1252 t   14 others
ORIGIN

US-10-045-180A-6 (1-31) x A98570 (1-4415)

Alignment Scores:
Pred. No.:   5.46e-18
Score: 191.00   Length: 4415
Percent Similarity: 100.00%   Matches: 31
Percent Identity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 6   Gaps: 0

Qy   1 IleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyValGlyThrCysPhe 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 4181 ATCTGCCATTGCAGAGTACTATACTGTGCATTTTGGAGAACATCTTGGTGGGACCTGCTTC 4240

Qy 21 IleLeuGlyGlyuArgTyrProIleCysCysTyr 31
  |||||||||||||||||||||||||||||||||
Db 4241 ATCCTTGGTGAACGCTACCCAATCTGCTGCTAC 4273

RESULT 4
BD074745
LOCUS BD074745 4415 bp DNA linear PAT 27-AUG-2002
DEFINITION Human defensin polypeptide Def-X, genome DNA and cDNA, composition
    containing the same, and application to diagnosis and remedy
    treatment.
ACCESSION BD074745
VERSION BD074745.1 GI:22620348
KEYWORDS JP 2001514264-A/1.
SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4415)
 AUTHORS Bougueleret.L. and Shmacov.I.
 TITLE Human defensin polypeptide Def-X, genome DNA and cDNA, composition
    containing the same, and application to diagnosis and remedy
 JOURNAL Patent: JP 2001514264-A 1 11-SEP-2001;
    GENSET
    OS Homo sapiens (human)
    PN JP 2001514264-A/1
    PD 11-SEP-2001
    PF 28-AUG-1998 JP 2000508701
    PR 29-AUG-1997 FR 97/10823
    PI LYDIE BOUGUELERET, ILIYA SHMACOV
    PC C07K14/435,A01N43/50,A01N63/00,A61K7/00,A61K38/00,A61P29/00,
    PC A61P35/00,
    PC A61P37/02,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N15/09, PC
    PC C12P21/02,
    PC C12P21/08,C12Q1/68,G01N33/53,A61K37/02,C12N15/00 CC
COMMENT Strandedness: Double;
    Topology: Linear;
    Human defensin polypeptide Def-X, genome DNA and cDNA, CC
    composition
    containing the same, and application to diagnosis and remedy
    treatment
FEATURES   Location/Qualifiers
 source 1..4415
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
 FH Key Location/Qualifiers
 FT exon 1836..1874
 FT exon 3394..3577
 FT exon 4161..4380
 FT CDS 3406..3408
 FT CDS 4276..4278.
BASE COUNT 1109 a 1128 c 912 g 1252 t 14 others
ORIGIN

US-10-045-180A-6 (1-31) x BD074745 (1-4415)

Alignment Scores:
Pred. No.:
Score: 5.46e-18
Percent Similarity: 191.00
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 6

     Length: 4415
     Matches: 31
     Conservative: 0
     Mismatches: 0
     Indels: 0
     Gaps: 0

Qy 1 IleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPhe 20
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4181 ATCTGCCATTGCAGAGTACTATACTGTGCATTTTGGAGAACATCTTGGTGGGACCTGCTTC 4240

Qy 21 IleLeuGlyGlyuArgTyrProIleCysCysTyr 31
  ||||||||||||||||||||||||||||
Db 4241 ATCCTTGGTGAACGCTACCCAATCTGCTGCTAC 4273

RESULT 5
AF238378/c
LOCUS AF238378 128544 bp DNA linear PRI 02-APR-2003
DEFINITION Homo sapiens chromosome 8 clone SCb-561b17 map p22-p21, complete
    sequence.
ACCESSION AF238378
VERSION AF238378.5 GI:29469504
KEYWORDS HTG.
SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    Chromosome 8 genomic sequence
REFERENCE 1 (bases 1 to 128544)
 AUTHORS Polley,A., Baumgart,C., Blechschmidt,K., Dette,M.D., Jahn,N.,
    Menzel,U., Reichwald,K., Schilhabel,M.B., Schudy,A., Tauden,S.,
    Wen,G., Schutte,B., Malik,M., Peng,J.Hong., McCray,P. and
    Rosenthal,A.
 TITLE Unpublished
 JOURNAL
REFERENCE 2 (bases 1 to 128544)
 AUTHORS Schudy,A., Schilhabel,M., Schutte,B., Ganz,T., Linzmeier,R.,
    Ho,C.H., Hoang,B.V., McCray,P., Baumgart,C., Menzel,U.,
    Schattevoy,R. and Rosenthal,A.
 TITLE Direct Submission
 JOURNAL Submitted (22-FEB-2000) Genome Analysis, Institute of Molecular
    Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE 3 (bases 1 to 128544)
 AUTHORS Genome Sequencing Center Jena.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular
    Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 4 (bases 1 to 128544)
 AUTHORS Genome Sequencing Center Jena.
 TITLE Direct Submission
 JOURNAL Submitted (10-NOV-2000) Genome Analysis, Institute of Molecular
    Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 5 (bases 1 to 128544)
 AUTHORS Lagemann,D. and Platzer,M.
 TITLE Direct Submission
 JOURNAL Submitted (06-JUL-2002) Genome Analysis, Institute of Molecular
    Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 6 (bases 1 to 128544)
 AUTHORS Lagemann,D. and Platzer,M.
 TITLE Direct Submission
 JOURNAL Submitted (02-APR-2003) Genome Analysis, Institute of Molecular
    Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT On Apr 2, 2003 this sequence version replaced gi:21700555.
    -------------------- Genome Center
    Center: Institute of Molecular Biotechnology
    Center code: IMB
    Web site: http://genome.imb-jena.de/
    Contact: gscj-submit@genome.imb-jena.de
    -------------------- Project Information
    Center project name: H370
    Center clone name: SCb-561b17
    -------------------- Summary Statistics
    Sequencing vector: pUCl8; 100% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 127999 bases at least Q40
    Consensus quality: 128330 bases at least Q30
    Consensus quality: 128544 bases at least Q20
    Quality coverage: 11.05x

    This sequence was finished as follows unless otherwise noted: all
    regions were double stranded, sequenced with an alternate
    chemistry, or covered by high quality data (i.e., phred quality >=
    30); an attempt was made to resolve all sequencing problems, such
    as compressions and repeats; all regions were covered by at least
    one plasmid subclone or more than one M13 subclone; and the
    assembly was confirmed by restriction digest.

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

```
FEATURES             Location/Qualifiers
source               1..128544
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="p22-p21"
                     /clone="SCb-561b17"
misc_feature         44529..44595
                     /note="single stranded/single chemistry region"
misc_feature         45703
                     /note="low quality region"
misc_feature         45815..45872
                     /note="low quality region"
misc_feature         45824..45872
                     /note="single stranded/single chemistry region"
unsure               45843
unsure               45847
unsure               45849
unsure               45858
unsure               45868
misc_feature         46235
                     /note="low quality region"
misc_feature         47899..47900
unsure               47899..47900
misc_feature         47172
                     /note="low quality region"
misc_feature         49172
                     /note="low quality region"
unsure               49172
misc_feature         49220
                     /note="low quality region"
misc_feature         49319
                     /note="low quality region"
unsure               49756..49807
misc_feature         49961
                     /note="single stranded/single chemistry region"
misc_feature         52933..52948
                     /note="low quality region"
misc_feature         52943
                     /note="low quality region"
unsure               52943
misc_feature         53544
                     /note="low quality region"
misc_feature         53612..53926
                     /note="low quality region"
misc_feature         53674..53677
                     /note="single stranded/single chemistry region"
misc_feature         53779..53799
                     /note="low quality region"
misc_feature         53787
                     /note="low quality region"
unsure               54734..55129
misc_feature         56028..56165
                     /note="single stranded/single chemistry region"
misc_feature         57952..58060
                     /note="single stranded/single chemistry region"
misc_feature         57968..57974
                     /note="low quality region"
unsure               57968..57974
unsure               57970
unsure               57974
misc_feature         58652
unsure               59561..59638
                     /note="low quality region"
misc_feature         59925
                     /note="single stranded/single chemistry region"
unsure               59929
misc_feature         61695..62103
                     /note="low quality region"
misc_feature         61858..61859
                     /note="single stranded/single chemistry region"
misc_feature         62263..62485
                     /note="low quality region"
misc_feature         62655..63118
                     /note="single stranded/single chemistry region"
misc_feature         62750..62785
                     /note="low quality region"
misc_feature         62774..62777
                     /note="low quality region"
unsure               62783
unsure               62801..62825
misc_feature         62819
misc_feature         62822
unsure               62872..62874
misc_feature         63119..63281
                     /note="low quality region"
misc_feature         62908
                     /note="low quality region"
misc_feature         62917..62987
                     /note="low quality region"
unsure               62933..62935
misc_feature         62955
unsure               62957
unsure               63082
misc_feature         63110
                     /note="low quality region"
unsure               63119..63141
misc_feature         63119..63141
                     /note="single clone coverage"
unsure               63129..63131
misc_feature         63142..63258
                     /note="single stranded/single chemistry region"
unsure               63159
misc_feature         63197..63258
unsure               63281
misc_feature         63876..63878
misc_feature         63933..64021
                     /note="low quality region"
misc_feature         63943
                     /note="single stranded/single chemistry region"
misc_feature         63984..63985
                     /note="low quality region"
misc_feature         64004..64007
                     /note="low quality region"
misc_feature         70360
                     /note="low quality region"
```

Alignment Scores:
Pred. No.:

US-10-045-180A-6 (1-31) x AF238378 (1-128544)

```
Qy        1 IleCysHisCysArgValLeuTyrCysIlePheGlyGluHisIleLeuGlyGlyThrCysPhe 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62550 ATCTGCCATTGCAGAGTACTATACTGCATTTTGGAGAACATCTTGGTGGGGACCTGGCTTC 62491

Qy       21 IleLeuGlyGluArgTyrProIleCysCysTyr 31
```

Qy 62490 ATCTTTGGTGAACGCTACCCAATCTGCTGCTAC 62458
         ||||||||||||||||||||||||||||||||
Db 62490 ATCTTTGGTGAACGCTACCCAATCTGCTGCTAC 62458

```
RESULT 6
LOCUS       AC116559
DEFINITION  AC116559   163093 bp    DNA     linear   HTG 30-MAY-2003
            Papio anubis clone rp41-339c10, WORKING DRAFT SEQUENCE, 3 ordered
            pieces.
ACCESSION   AC116559
VERSION     AC116559.18  GI:31192968
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Papio anubis (olive baboon)
ORGANISM    Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 163093)
  AUTHORS   Xu,W., Hua,A., Eichler,E. and Roe,B.A.
  TITLE     Papio anubis BAC Clone rp41-339c10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163093)
  AUTHORS   Xu,W., Hua,A., Eichler,E. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 163093)
  AUTHORS   Xu,W., Hua,A., Eichler,E. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma
            Center code:UOKNOR
COMMENT     On May 30, 2003 this sequence version replaced gi:30725962.
            Center: Department Of Chemistry And Biochemistry
            Center: Genome Center
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *     1: contig of 6315 bp in length
            *  6315: gap of unknown length
            *  6316: contig of 10289 bp in length
            *  6416: contig of 16704 bp in length
            * 16705: gap of unknown length
            * 16805: contig of 146289 bp in length.
FEATURES             Location/Qualifiers
     source          1..163093
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="rp41-339c10"
                     /clone_lib="RPCI - 41 Male (Olive) Baboon BAC Library"
BASE COUNT    44435 a   36508 c   35861 g   46065 t    224 others
ORIGIN
```

Alignment Scores:
Pred. No.:   1
Score: 1.08e-11
Percent Similarity: 159.00   Length: 163093
Best Local Similarity: 90.00%   Matches: 26
Query Match: 86.67%   Conservative: 1
DB: 83.25%   Mismatches: 3
                         2   Indels: 0
                                         Gaps: 0

US-10-045-180A-6 (1-31) x AC116559 (1-163093)

Qy 2 CysHisCysArgValLeuTyrCysIlePheGlyHisLeuGlyGlyThrCysPheIle 21

Qy 22 LeuGlyGluArgTyrProIleCysCysTyr 31
   33641 TGCCAITGCAGAGTACTTCTPACTGCCCTTTGGGAGGAACATCTTTGGTGGGAACTTCCTTCATC 33700
         ||||||||||||||||||||||||||||||||::::||||||||||||||||||||||||||||
Db 33701 CATGGTGAACGCTCCCAATCTGCTGCTAC 33730

```
RESULT 7
LOCUS       AC116558/c
DEFINITION  AC116558   188604 bp    DNA     linear   HTG 11-JUN-2003
            Papio anubis clone rp41-273g19, WORKING DRAFT SEQUENCE, 2 ordered
            pieces.
ACCESSION   AC116558
VERSION     AC116558.16  GI:30725961
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Papio anubis (olive baboon)
ORGANISM    Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 188604)
  AUTHORS   Xu,W., Hua,A., Eichler,E. and Roe,B.A.
  TITLE     Papio anubis BAC Clone rp41-273g19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188604)
  AUTHORS   Xu,W., Hua,A., Eichler,E. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 188604)
  AUTHORS   Xu,W., Hua,A., Eichler,E. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 15, 2003 this sequence version replaced gi:28173120.
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            Center: Genome Center
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *     1: contig of 6057 bp in length
            *  6057: gap of unknown length
            *  6058: contig of 182447 bp in length
            *  6157: gap of unknown length
            *  6158: contig of 182447 bp in length.
FEATURES             Location/Qualifiers
     source          1..188604
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="rp41-273g19"
                     /clone_lib="RPCI - 41 Male (Olive) Baboon BAC Library"
BASE COUNT    53384 a   41769 c   40796 g   52555 t    100 others
ORIGIN
```

Alignment Scores:
Pred. No.:   1
Score: 1.25e-11
Percent Similarity: 159.00   Length: 188604
Best Local Similarity: 90.00%   Matches: 26
Query Match: 86.67%   Conservative: 1
DB: 83.25%   Mismatches: 3
                         2   Indels: 0
                                         Gaps: 0

US-10-045-180A-6 (1-31) x AC116558 (1-188604)

Qy 2 CysHisCysArgValLeuTyrCysIlePheGlyHisLeuGlyGlyThrCysPheIle 21

```
Qy    2 CysHisCysArgValLeuTyrCysIlePheGlyGlyGluHisLeuGlyThrCysPheIle 21
        |||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db 141891 TGCCATGCAGGAGTACTCTACTGCCCTTTTGGAGAGACATCTTGGTGGGACCTTCCTTCATC 141832

Qy   22 LeuGlyGluArgTyrProIleCysCysTyr 31
        ||||||||||||||||||||||||||||||
Db 141831 CATGGTGAACGCTCCCCAATCTGCTGCTAC 141802
```

RESULT 8
```
AC098442/2
AC098442/c
LOCUS       AC098442            271979 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-2J1, *** SEQUENCING IN PROGRESS ***,
            3 unordered pieces.
ACCESSION   AC098442
VERSION     AC098442.5  GI:24941300
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 271979)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davy-Carroll,L., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,F., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A.V., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,D., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,C., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wieczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 271979)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 271979)
  AUTHORS   Worley,K.C.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:23664583.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            --------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            --------------- Project Information
            Center project name: TUSW
            Center clone name: CH230-2J1
            --------------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 236907 bases at least Q40
            Consensus quality: 240990 bases at least Q30
            Consensus quality: 243222 bases at least Q20
            Estimated insert size: 243581; sum-of-contigs estimation
            Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            * NOTE: Estimated insert size may differ from sequence length
            *    (see http://www.hgsc.bcm.tmc.edu/docs/Genbank-draft-data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1..271979
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-2J1"
     misc_feature    1..1234
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    268032..268032: contig of 268032 bp in length
     misc_feature    268033..268132: gap of unknown length
     misc_feature    268133..270509: contig of 2377 bp in length
     misc_feature    270509..270609: gap of unknown length
     misc_feature    270610..271979: contig of 1370 bp in length.
     misc_feature    2826..4181
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    9192..9978
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    150206..150798
                     /note="clone_boundary
                     clone_end:Sp6"
                     end_sequence:BH288135"
     misc_feature    150206..150798
                     end_sequence:BH288134"
                     /note="clone_boundary
                     clone_end:T7"
     misc_feature    266939..268032
                     site:
                     end_sequence:BH288134"
                     /note="wgs_end_extension
                     clone_end:T7"
```

```
BASE COUNT     74463 a   48993 c   50388 g   70500 t   27635 others
ORIGIN
```

```
US-10-045-180A-6 (1-31) x AC098442 (1-271979)

Qy          2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuArgGlyGlyThrCysPheIle 21
              |||::::|||||||||||:||||:|||||:|||||||||||||||||||||||||:|||||
Db     228016 TGCTATTGCAGACTCCTCAGCTGTCAATTTGGGGACAGACTCGCTGGGTCCTGCCGTTCT 227957

Qy         22 LeuGlyGluArgTyrProIleCysCysTyr 31
              ||:|||||||||||||||:|||||::|:
Db     227956 GGTGGCGGTCAACTACCCCTTCGTCGTGCCAC 227927
```

```
RESULT 9
LOCUS       AC128185               301130 bp    DNA     linear   HTG 22-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-30N20, ***  SEQUENCING IN PROGRESS
            ***, 4 unordered pieces.
ACCESSION   AC128185
VERSION     AC128185.2  GI:22367661
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 301130)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,J., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,J.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,D.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,L., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarinpunagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,J., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,J., Reeves,C., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P.,
            Rives,C., Rodkey,T., Rojas,A., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Scott,G., Shatsman,S., Shen,H.,
```

```
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soza,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,J., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wieczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,P., Zhang,J., Zhou,X., Zhao,S., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 301130)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 301130)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 22, 2002 this sequence version replaced gi:21908787.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            --------------  Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            --------------  Project Information
            Center project name: GYBD
            Center clone name: CH230-30N20
            --------------  Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 240138 bases at least Q40
            Consensus quality: 244747 bases at least Q30
            Consensus quality: 244371 bases at least Q20
            Estimated insert size: 272390; sum-of-contigs estimation
            Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *     1      3552:  contig of 3552 bp in length
            *  3553      3552:  gap of unknown length
            *  3553     16658:  contig of 13006 bp in length
            * 16658     16759:  gap of unknown length
            * 16759     29649:  contig of 279732 bp in length
            * 296491    296590:  gap of unknown length
            * 296591   301130:  contig of 4540 bp in length.
FEATURES             Location/Qualifiers
     source          1..301130
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-30N20"
```

```
FEATURES
     misc_feature    1..1039
                     /note="wgs end extension
                     clone end:Sp6"
     misc_feature    1631..3552
                     /note="wgs end extension
                     clone end:Sp6"
     misc_feature    3653..4875
                     /note="wgs end extension
                     clone end:Sp6"
     misc_feature    complement(8004..8619)
                     /note="wgs end extension
                     clone end:Sp6"
     misc_feature    /note="clone boundary
                     site:EcoRI
                     end sequence:BH293755"
     misc_feature    12505..13006
                     /note="clone boundary
                     site:EcoRI
                     end sequence:BH293754"
                     277504..279612
                     /note="wgs end extension
                     clone end:T7"

BASE COUNT    74053 a   51646 c   49686 g   73702 t   52043 others
ORIGIN

US-10-045-180A-6 (1-31) x AC128185 (1-301130)

Alignment Scores:
Pred. No.:              0.00776        Length:       301130
Score:                101.00           Matches:      16
Percent Similarity:    70.00%          Conservative: 5
Best Local Similarity: 53.33%          Mismatches:   9
Query Match:           52.88%          Indels:       0
DB:                    2               Gaps:         0

Qy        2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
            ||:||::|||:|:|||:::||::|||:|||||||||:::||
Db    44227 TGCTATTGCAGAGCTCCTCAGCGTGTCAATTTGGGGAACGACTCGGCTTATGGGTGGGGA 44286

Qy       22 LeuGlyGluArgTyrProIleCysCysTyr 31
            |||::||||:||
Db    44287 GGTGGGCGTCACCTACCCCCTCTGCTGCCAC 44316


RESULT 10
LOCUS       AF188268            347 bp    mRNA    linear   PRI 07-NOV-1999
DEFINITION  Macaca mulatta defensin-1 mRNA, complete cds.
ACCESSION   AF188268
VERSION     AF188268.1  GI:6273105
KEYWORDS    .
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 347)
  AUTHORS   Tang,Y.Q., Yuan,J., Miller,C.J. and Selsted,M.E.
  TITLE     Isolation, characterization, cDNA cloning, and antimicrobial
            properties of two distinct subfamilies of alpha-defensins from
            rhesus macaque leukocytes
  JOURNAL   Infect. Immun. 67 (11), 6139-6144 (1999)
  MEDLINE   20002603
   PUBMED   10531277
REFERENCE   2  (bases 1 to 347)
  AUTHORS   Selsted,M.E.
  TITLE     Rhesus myeloid defensins
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 347)
  AUTHORS   Selsted,M.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-1999) Pathology, UC Irvine, College of Medicine,
            Irvine, CA 92697, USA
FEATURES             Location/Qualifiers
     source          1..347
                     /organism="Macaca mulatta"
                     /mol_type="mRNA"
                     /db_xref="taxon:9544"
                     /tissue_type="myeloid"
     CDS             52..342
                     /codon_start=1
                     /product="defensin-1"
                     /protein_id="AAF06312.1"
                     /db_xref="GI:6273106"
                     /translation="MRTLVILAAILLVALQAQAEPLQARTDEATAAQEQIPTDNPEVV
                     VSLAWDESLAPKDSVPGLRRKNMACYCRIPACLAGERRYGTCFYLGRVWAFCC"
BASE COUNT     76 a    102 c     94 g     75 t
ORIGIN

US-10-045-180A-6 (1-31) x AF188268 (1-347)

Alignment Scores:
Pred. No.:              0.000565       Length:       347
Score:                 89.00          Matches:      14
Percent Similarity:    62.07%          Conservative: 4
Best Local Similarity: 48.28%          Mismatches:   11
Query Match:           46.60%          Indels:       0
DB:                    9               Gaps:         0

Qy        2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
            ||:||::||||:::||::|||||||||||:||||||
Db      253 TGCTATTGCAGAATACCAGCGTGTCCTTAGCAGGAGGAAGAACGTTCGGCTTATGGAACC 312

Qy       22 LeuGlyGluArgTyrProIleCysCys 30
            :::||||||
Db      313 CTCGGCAAGAGTCTGGGCATTCTGCTGC 339


RESULT 11
LOCUS       AF184160            435 bp    mRNA    linear   PRI 14-NOV-1999
DEFINITION  Macaca mulatta alpha-defensin 1A (MNP1A) mRNA, complete cds.
ACCESSION   AF184160
VERSION     AF184160.1  GI:6409289
KEYWORDS    .
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   Zhao,C., Nguyen,T. and Lehrer,R.I.
  TITLE     cDNA cloning of three alpha-defensins and three demidefensins from
            rhesus monkey bone marrow
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 435)
  AUTHORS   Zhao,C., Nguyen,T. and Lehrer,R.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1999) Medicine, UCLA, 10833 le Conte Avenue, Los
            Angeles, CA 90095, USA
FEATURES             Location/Qualifiers
     source          1..435
                     /organism="Macaca mulatta"
                     /mol_type="mRNA"
                     /db_xref="taxon:9544"
     gene            13..303
                     /gene="MNP1A"
     CDS             13..303
                     /gene="MNP1A"
                     /codon_start=1
                     /product="alpha-defensin 1A"
                     /protein_id="AAF07926.1"
                     /db_xref="GI:6409290"
                     /translation="MRTLAILAAILLVALQAQAEPLQARTDEATAAQEQIPTDNPRVV
                     VSLAWDESLAPKDSVPGLRRKNMACYCRIPACLAGERRYGTCFYLGRVWAFCC"
BASE COUNT    120 a    113 c     104 g     98 t
ORIGIN
```

RESULT 12

Alignment Scores:
Pred. No.: 2

Score:	89.00	
Percent Similarity:	62.07%	Length: 435
Best Local Similarity:	48.28%	Matches: 14
Query Match:	46.60%	Conservative: 4
		Mismatches: 11
		Indels: 0
		Gaps: 0

US-10-045-180a-6 (1-31) x AF184160 (1-435)

```
Qy    22  LeuGlyGluArgTyrProIleCysCys  30
          |||||||||||||||||||||||||||
Db   274  CTGGGAAGAGTCTGGGCATTCTGCTGC  300
```

RESULT 12

LOCUS   AF184159   433 bp   mRNA   linear   PRI 14-NOV-1999
DEFINITION Macaca mulatta alpha-defensin 1 (MNP1) mRNA, complete cds.
ACCESSION   AF184159
VERSION   AF184159.1 GI:6409287
KEYWORDS   .
SOURCE   Macaca mulatta (rhesus monkey)
ORGANISM   Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1 (bases 1 to 433)
  AUTHORS   Zhao,C., Nguyen,T. and Lehrer,R.I.
  TITLE   cDNA cloning of three alpha-defensins and three demidefensins from
            rhesus monkey bone marrow
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 433)
  AUTHORS   Zhao,C., Nguyen,T. and Lehrer,R.I.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-SEP-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los
            Angeles, CA 90095, USA
FEATURES   Location/Qualifiers
     source   1..433
                     /organism="Macaca mulatta"
                     /mol_type="mRNA"
                     /db_xref="taxon:9544"
     gene   1..433
                     /gene="MNP1"
     CDS   13..303
                     /gene="MNP1"
                     /codon_start=1
                     /product="alpha-defensin 1"
                     /protein_id="AAF07925.1"
                     /db_xref="GI:6409288"
                     /translation="MRTLAILAAILLVALQAQAEPLQARTDEATAAQEQIPTDNPEVV
                     VSLAWDESLAPKDSVPGLRKNMACYCRIPACLAGERRYGTCFYMGRVWAFCC"
BASE COUNT   121 a   113 c   102 g   97 t
ORIGIN

US-10-045-180a-6 (1-31) x AF184159 (1-433)

Alignment Scores:
Pred. No.: 2

Score:	87.00	
Percent Similarity:	62.07%	Length: 433
Best Local Similarity:	44.83%	Matches: 13
Query Match:	45.55%	Conservative: 5
		Mismatches: 11
		Indels: 0
		Gaps: 0

```
Qy     2  CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle  21
          ||:||||||||  :::
Db   214  TGCTATTGCAGAATACCAGCGGTGCTTAGCAGGAAGCGTTCGGCTTATGGAAGAGAACCTGCTTCTAC  273

Qy    22  LeuGlyGluArgTyrProIleCysCys  30
```

Db   274 ATCGGAAGAGTCTGGGCATTCTGCTGC 300

RESULT 13

LOCUS   RNU50356   1260 bp   DNA   linear   ROD 05-JAN-1999
DEFINITION Rattus norvegicus neutrophil defensin pseudogene, partial cds.
ACCESSION   U50356
VERSION   U50356.1 GI:4090429
KEYWORDS   .
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 1260)
  AUTHORS   Banaiee,N., Yount,N.Y. and Selsted,M.E.
  TITLE   Molecular Characterization of Genes Encoding Rat Neutrophil
            Defensins
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1260)
  AUTHORS   Banaiee,N., Yount,N.Y. and Selsted,M.E.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-FEB-1996) Niaz Banaiee, Pathology, UC Irvine, D440
            Med. Sci., Irvine, CA 92717, USA
FEATURES   Location/Qualifiers
     source   1..1260
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
     repeat_region   <1..171
                     /number=1
     intron   <1..171
                     /number=1
     repeat_region   66..126
                     /note="CA dinucleotide variable number repeat element"
     CDS   join(185..356,861..970)
                     /pseudo
                     /codon_start=1
                     /product="neutrophil defensin"
     exon   172..356
                     /number=2
     intron   357..860
                     /number=2
     exon   861..1096
                     /number=3
     intron   978..989
     repeat_region   978..989
                     /note="GAA trinucleotide variable number repeat element"
BASE COUNT   372 a   271 c   243 g   374 t
ORIGIN

US-10-045-180A-6 (1-31) x RNU50356 (1-1260)

Alignment Scores:
Pred. No.: 2

Score:	85.50	
Percent Similarity:	65.52%	Length: 1260
Best Local Similarity:	51.72%	Matches: 15
Query Match:	44.76%	Conservative: 4
		Mismatches: 9
		Indels: 1
		Gaps: 1

```
Qy     2  CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle  21
          ||:||||||||  :::
Db   881  TGCTATTGCAGAATCC--AGCTGTCAATTTGGGGAACGACTCCGCGGGTCCGTGCCGTTCC  937

Qy    22  LeuGlyGluArgTyrProIleCysCys  30
          |||:|||
Db   938  GGTGGCATCACCTACCCCTCTGCTGC  964
```

RESULT 14

LOCUS   RNU16684   494 bp   mRNA   linear   ROD 12-MAR-1996
DEFINITION Rattus norvegicus defensin RatNP-4 precursor mRNA, complete cds.
ACCESSION   U16684

```
VERSION      U16684.1  GI:1041806
KEYWORDS     .
SOURCE       Rattus norvegicus (Norway rat)
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 494)
  AUTHORS    Yount,N.Y., Wang,M.S., Yuan,J., Banaiee,N., Ouellette,A.J. and
             Selsted,M.E.
  TITLE      Rat neutrophil defensins. Precursor structures and expression
             during neutrophilic myelopoiesis
  JOURNAL    J. Immunol. 155 (9), 4476-4484 (1995)
  MEDLINE    96025910
   PUBMED    7594610
REFERENCE    2  (bases 1 to 494)
  AUTHORS    Yount,N.Y.
  TITLE      Direct Submission
  JOURNAL    Submitted (27-OCT-1994) Nannette Y. Yount, Pathology, University of
             California, D440 Med. Sci. I, Irvine, CA 92717, USA
FEATURES             Location/Qualifiers
     source          1..494
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /tissue_type="bone marrow"
                     /clone_lib="plasmid, oligo dT primed"
     CDS             72..353
                     /codon_start=1
                     /product="defensin RatNP-4 precursor"
                     /protein_id="AAA91972.1"
                     /db_xref="GI:1041807"
                     /translation="MRITLTLLTFLILLALHTQAESPQERAKAAPDQDMVMEDQDIPIS
                     FGGYKGTVLQDAVVKAGQACYCRIGACVSGERLTGACLNGRIYRLCCR"
     sig_peptide     72..128
     mat_peptide     258..350
                     /product="defensin peptide RatNP-4"
     repeat_region   358..400
                     /note="Repeat element is not perfect.  In six clones
                     analyzed repeat element is identical"
     polyA_signal    478..483
     polyA_site      494
BASE COUNT      123 a    119 c    129 g    123 t
ORIGIN
```

Alignment Scores:
```
Pred. No.:              0.00442        Length:        494
Score:                  84.00          Matches:        13
Percent Similarity:     62.07%         Conservative:    5
Best Local Similarity:  44.83%         Mismatches:     11
Query Match:            43.98%         Indels:          0
                                       Gaps:            0
```

US-10-045-180A-6 (1-31) x RNU16684 (1-494)

```
Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGlyTyrLysGlyThrValLeuGlnAspAlaValValLysAlaGlyGlnAlaCysTyrCysArgIleGlyAlaCysPheIle 21
         ||| :||||| |||   ::|||||||||
Db 261 TGCTATTGCAGAATCGGAGCCTGTGTTTCTGGAGAACCGGCTCACTGGGGCGCATGGGTCTCGTC 320

Qy  22 LeuGlyGluArgTyrProIleCysCys 30
       ||| :::|||||||
Db 321 AATGGCCCGCATCTACCGTCTCGTTGC 347
```

```
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 504)
  AUTHORS    Yount,N.Y., Wang,M.S., Yuan,J., Banaiee,N., Ouellette,A.J. and
             Selsted,M.E.
  TITLE      Rat neutrophil defensins. Precursor structures and expression
             during neutrophilic myelopoiesis
  JOURNAL    J. Immunol. 155 (9), 4476-4484 (1995)
  MEDLINE    96025910
   PUBMED    7594610
REFERENCE    2  (bases 1 to 504)
  AUTHORS    Yount,N.Y.
  TITLE      Direct Submission
  JOURNAL    Submitted (27-OCT-1994) Nannette Y. Yount, Pathology, University of
             California, D440 Med. Sci. I, Irvine, CA 92717, USA
FEATURES             Location/Qualifiers
     source          1..504
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /tissue_type="bone marrow"
                     /clone_lib="plasmid, oligo dT primed"
     CDS             62..325
                     /codon_start=1
                     /product="defensin RatNP-3 precursor"
                     /protein_id="AAA91971.1"
                     /db_xref="GI:1041805"
                     /translation="MRITLTLLTFLILLALHTQAESPQGSTKEAPDEEQDISVFFGGDK
                     GTALQDDAAVKAGVTCSCRTSSCRFGERLSGACLNGRIYRLCC"
     sig_peptide     62..118
     mat_peptide     236..322
                     /product="defensin peptide RatNP-3"
     repeat_region   333..404
                     /note="cDNA sequence shown with 24 repeats of the
                     polypurine GAA sequence.  In an additional two clones the
                     GAA-motif was repeated 28 times and in one clone it was
                     repeated 31 times"
     polyA_signal    483..488
     polyA_site      504
BASE COUNT      149 a    117 c    121 g    117 t
ORIGIN
```

Alignment Scores:
```
Pred. No.:              0.0045         Length:        504
Score:                  84.00          Matches:        14
Percent Similarity:     55.17%         Conservative:    2
Best Local Similarity:  48.28%         Mismatches:     13
Query Match:            43.98%         Indels:          0
                                       Gaps:            0
```

US-10-045-180A-6 (1-31) x RNU16683 (1-504)

```
Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGlyHisIleLeuGlyTyrThrCysSerPheIle 21
         || ||  |||||        |||||||
Db 236 TGCTCCTTGTAGAACCTCAAGCTGTCGTTTGGAGAAGAGCCTCTGGGGCGGCATGTCGTCTCTC 295

Qy  22 LeuGlyGluArgTyrProIleCysCys 30
       ||  :::||||||||
Db 296 AATCGCCCGCATCTACCGACTCTGTTGC 322
```

US-10-045-180A-6 (1-31) x RNU16683 (1-504)

RESULT 15
```
LOCUS       RNU16683     504 bp    mRNA    linear   ROD 12-MAR-1996
DEFINITION  Rattus norvegicus defensin RatNP-3 precursor mRNA, complete cds.
ACCESSION   U16683
VERSION     U16683.1  GI:1041804
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
```

Search completed: December 17, 2003, 16:40:23
Job time : 961.963 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 14:50:33 ; Search time 1083.85 Seconds
                 (without alignments)
                 695.153 Million cell updates/sec

Title: US-10-045-180A-6
Perfect score: 191
Sequence: 1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31

Scoring table: BLOSUM62
                 Xgapop 10.0 , Xgapext 0.5
                 Ygapop 10.0 , Ygapext 0.5
                 Fgapop 6.0 , Fgapext 7.0
                 Delop   6.0 , Delext 7.0

Searched: 22781392 seqs, 1215223056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat_17122003_145024_12540/app_query.fasta_1.860
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DQCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10045180 @CGN_1 1 6100 @runat_17122003_145024_12540 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
                  1: em_estba:*
                  2: em_esthum:*
                  3: em_estin:*
                  4: em_estmu:*
                  5: em_estvr:*
                  6: em_estpl:*
                  7: em_estro:*
                  8: em_htc:*
                  9: em_est1:*
                 10: gb_est2:*
                 11: gb_htc:*
                 12: gb_est3:*
                 13: gb_est4:*
                 14: gb_est5:*
                 15: em_estfun:*
                 16: em_esthum:*
                 17: em_estin:*
                 18: em_gss_hum:*
                 19: em_gss_inv:*
                 20: em_gss_pln:*
                 21: em_gss_fun:*
                 22: em_gss_mam:*
                 23: em_gss_mus:*
                 24: em_gss_pro:*
                 25: em_gss_rod:*
                 26: em_gss_phg:*
                 27: em_gss_vrl:*
                 28: gb_gss1:*
                 29: gb_gss2:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.		Pred. Score	% Query Match	Match Length	DB	ID	Description
1		164	85.9	28	606	AQ624584	AQ624584 HS_2110_B
2	c	88	46.1	10	329	BF904561	BF904561 PM3-MT020
3	c	84	44.0	9	454	AI403357	AI403357 EST2366647
4	c	84	44.0	10	468	BF290038	BF290038 EST454629
5	c	84	44.0	13	547	BQ200749	BQ200749 UI-R-RB1-
6	c	82	42.9	13	287	AW467913	AW467913 he28d03.x
7	c	82	42.9	12	349	BI021835	BI021835 RC5-MT025
8	c	80	41.9	13	181	BQ334373	BQ334373 QV1-MT016
9	c	80	41.9	13	219	BE831361	BE831361 QV4-MT005
10		80	41.9	9	225	AA321199	AA321199 EST24002
11		80	41.9	9	229	AA321481	AA321481 EST23961
12	c	80	41.9	9	233	AW057767	AW057767 xa85b06.x
13	c	80	41.9	12	234	BF893374	BF893374 QV3-MT012
14	c	80	41.9	12	241	BI028909	BI028909 PM3-MT020
15	c	80	41.9	10	246	BF903426	BF903426 IL2-MT017
16	c	80	41.9	12	251	BI023343	BI023343 CM4-MT024
17	c	80	41.9	10	255	BE831399	BE831399 QV4-MT005
18	c	80	41.9	10	257	BE831404	BE831404 QV4-MT005
19	c	80	41.9	9	259	AW274523	AW274523 xv30h12.x
20	c	80	41.9	10	260	BE831426	BE831426 QV4-MT005
21	c	80	41.9	12	269	BI022826	BI022826 CM0-MT034
22	c	80	41.9	10	273	BE831401	BE831401 QV4-MT005
23	c	80	41.9	10	274	BF901251	BF901251 IL2-MT017
24	c	80	41.9	10	274	BF901407	BF901407 IL2-MT017
25	c	80	41.9	10	274	BF901500	BF901500 IL2-MT017
26	c	80	41.9	10	274	BF903483	BF903483 IL2-MT017
27	c	80	41.9	9	274	BF905787	BF905787 IL2-MT017
28	c	80	41.9	9	279	AI435021	AI435021 th76h11.x
29	c	80	41.9	13	286	BQ344909	BQ344909 IL2-MT017
30	c	80	41.9	10	288	BE831389	BE831389 QV4-MT005
31	c	80	41.9	10	289	BE831385	BE831385 QV4-MT005
32	c	80	41.9	10	290	BE831314	BE831314 QV4-MT004
33	c	80	41.9	10	290	BE831369	BE831369 QV4-MT005
34	c	80	41.9	10	290	BE831382	BE831382 QV4-MT005
35	c	80	41.9	10	290	BE831396	BE831396 QV4-MT005
36	c	80	41.9	10	290	BE831397	BE831397 QV4-MT005
37	c	80	41.9	10	290	BE831398	BE831398 QV4-MT005
38	c	80	41.9	10	290	BE831402	BE831402 QV4-MT005
39	c	80	41.9	10	290	BE831405	BE831405 QV4-MT005
40	c	80	41.9	10	290	BE831406	BE831406 QV4-MT005
41	c	80	41.9	10	290	BE831407	BE831407 QV4-MT005
42		80	41.9	10	290	BE831411	BE831411 QV4-MT005
43		80	41.9	10	291	BE831414	BE831414 QV4-MT005
44		80	41.9	10	291	BE831311	BE831311 QV4-MT004
45		80	41.9	10	298	BE831364	BE831364 QV4-MT005

ALIGNMENTS

RESULT 1
AQ624584/c
LOCUS   AQ624584   606 bp   DNA   linear   GSS 16-JUN-1999
DEFINITION HS_2110_B1_D06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2110 Col=11 Row=H, genomic survey
            sequence.
ACCESSION   AQ624584
VERSION   AQ624584.1 GI:5086976
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 606)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE   Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
PUBMED   10449764
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
FEATURES   Location/Qualifiers
     source   1..606
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /sex="male"
                     /clone="plate=2110 Col=11 Row=H"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
                     Plate: 2110 row: H column: 11
                     Seq primer: M13 Reverse
                     Class: BAC ends
                     High quality sequence stop: 606.
BASE COUNT   169 a   128 c   131 g   172 t   6 others
ORIGIN

US-10-045-180A-6 (1-31) x AQ624584 (1-606)

Alignment Scores:
Pred. No.:   1. ..606
Score: 2.59e-11
Percent Similarity: 164.00
Best Local Similarity: 90.32%
Query Match: 90.32%
                    85.86%
DB: 28

Qy   1 IleCysHisCysArgValLeuTyrCysIlePheGlyGlyHisLeuGlyGlyThrCysPhe 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150 ATCTGCCATTGCAGAGTACTACTACGTCATTTNTGGAGAACTCTTGGTAGGACAAAATTC 91

Qy   21 IleLeuGlyGlyHisArgTyrProIleCysCysTyr 31
         ||||||||||||||||||||||||||||||
Db   90 ATCCTTGGTGAACGCTACCCCATCTGCTGCTAC 58

JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
            Tel: +55-11-27704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM3&t2=PM3-MT0200-211200-002-a11&t3=2000-12-21&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 2.
            High quality sequence stop: 329.
FEATURES   Location/Qualifiers
     source   1..329
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0200"
                     /note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT   76 a   82 c   85 g   85 t   1 others
ORIGIN

US-10-045-180A-6 (1-31) x BF904561 (1-329)

Alignment Scores:
Pred. No.:   2. ..329
Score: 0.0852
Percent Similarity: 88.00
Best Local Similarity: 58.62%
Query Match: 58.62%
                    48.28%
DB: 10

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGlyHisLeuGlyGlyThrCysPheIle 21
         |||||:::||||||||||||||||||||||||||||||||||||||
Db 194 TGCTATTGCAGAATACCAGCCGTGCATTGCAGGGAGAACGTCGGCTARGGAACCTGCATCTAC 135

Qy   22 LeuGlyGlyHisArgTyrProIleCysCys 30
         ::|||||||||||::||||||
Db 134 CAGGGAGAGACTTCTGGGCCATTCTGCTGC 108

RESULT 2
LOCUS   BF904561
DEFINITION PM3-MT0200-211200-002-a11 MT0200 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF904561
VERSION   BF904561.1 GI:12296020
KEYWORDS   EST.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 329)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare, M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE   Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

---

RESULT 3
LOCUS   AI408357/c
DEFINITION EST236647 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone ROVP882 3' end, mRNA sequence.
ACCESSION   AI408357
VERSION   AI408357.1 GI:4251861
KEYWORDS   EST.
SOURCE   Rattus sp.
  ORGANISM Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 454)
AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
TITLE   Rat Gene Index
            Gene Index
            Unpublished
JOURNAL   Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index
COMMENT   Contact: Lee, NH

LOCUS   AI408357/c   454 bp   mRNA   linear   EST 09-FEB-1999

Left column

```
                The Institute for Genomic Research
                9712, Medical Center Drive, Rockville, MD 20850, USA
                Tel: (301)-838-3529
                Fax: (301)-838-0208
                Email: nhlee@tigr.org
                Seq primer: M13-21.

FEATURES        Location/Qualifiers
     source     1. .454
                /organism="Rattus sp."
                /mol_type="mRNA"
                /db_xref="taxon:10118"
                /clone="ROVEP82"
                /note="Normalized rat ovary, Bento Soares"
                /clone_lib="Normalized rat ovary, Bento Soares"
                /note="Organi: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
                Site_2: NotI"

BASE COUNT      111 a    118 c    106 g    119 t
ORIGIN
```

US-10-045-180A-6 (1-31) x AI408357 (1-454)

Alignment Scores:
```
Pred. No.:              0.434
Score:                  84.00
Percent Similarity:     62.07%
Best Local Similarity:  44.83%
Query Match:            43.98%
DB:                     9
```

```
Qy    2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
        |||:::|||||:::|::||
Db  239 TGCTATTGCAGAAATCGGAGCCCTGTGTTTCTGGAGAACGGCTCACTGGGCGCATGTGGTCTC 180

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
        ||||:||||||
Db  179 AATGGCCGCATCTACCGTCTCTGTTGC 153
```

```
RESULT 4
LOCUS       BF290038          468 bp    mRNA    linear    EST 28-NOV-2000
DEFINITION  EST454629 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
            clone RGIHM93 3' sequence, mRNA sequence.
ACCESSION   BF290038
VERSION     BF290038.1  GI:11121108
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,B., Pertea,G.,
            Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
  TITLE     Generation of ESTs from Normalized Rat Embryo, Bento Soares
  JOURNAL   Unpublished
COMMENT     Other_ESTs: EST352370
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            This clone is available through the ATCC, contact the ATCC
            for further information.
            tel#703-365-2700
FEATURES    Location/Qualifiers
     source 1. .468
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /db_xref="taxon:10116"
            /clone="RGIHM93"
            /tissue_type="mixed tissue"
            /lab_host="DH5-alpha"
            /clone_lib="Rat Gene Index, normalized rat, Rattus
            norvegicus cDNA"
```

Right column

/note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROV, RBR, RKI, RLI, RPL, RLU, REM, RMU, RSP,
RHE, RPC, RPN"

```
BASE COUNT      117 a    122 c    110 g    119 t
ORIGIN
```

US-10-045-180a-6 (1-31) x BF290038 (1-468)

Alignment Scores:
```
Pred. No.:              0.452
Score:                  84.00
Percent Similarity:     62.07%
Best Local Similarity:  44.83%
Query Match:            43.98%
DB:                     10
```

```
Qy    2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
        |||:::|||||:::|::||
Db  239 TGCTATTGCAGAATCGGAGCCCTGTGTTTCTGGAGAACGGCTCACTGGGCGCATGTGGTCTC 180

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
        ||||:||||||
Db  179 AATGGCCGCATCTACCGTCTCTGTTGC 153
```

```
RESULT 5
LOCUS       BQ200749          547 bp    mRNA    linear    EST 02-MAY-2002
DEFINITION  UI-R-EB1-clq-e-09-0-UI.s1 UI-R-EB1 Rattus norvegicus cDNA clone
            UI-R-EB1-clq-e-09-0-UI 3', mRNA sequence.
ACCESSION   BQ200749
VERSION     BQ200749.1  GI:20417214
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized duodenum library cDNA Library Preparation: M.B. Soares
            lab Clone distribution: clones will be available through Research
            Genetics (www.resgen.com) The following repetitive elements were
            found in this cDNA sequence: 118-200, >(GAA)n#simple_repeat
            POLYA=Yes.
FEATURES    Location/Qualifiers
     source 1. .547
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /strain="Sprague-Dawley"
            /db_xref="taxon:10116"
            /clone="UI-R-EB1-clq-e-09-0-UI"
            /dev_stage="adult"
            /lab_host="DH10B (Life Technologies)"
            /clone_lib="UI-R-EB1"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
```

Left column

```
                    polylinker; Site_1: Not I; Site_2: Eco RI; UI-R-EB1 is a
                    non-normalized Rat Duodenum library (RDU) constructed in
                    pT3T7 PAC vector according to the procedure described by
                    Bonaldo, Lennon & Soares (Genome Research Genome 6:
                    791-806, 1996). The oligonucleotide used to prime first
                    strand synthesis contained the sequence tag TGTGGTTCAT
                    between the Not I cloning site and dT18 stretch. The Rat
                    Duodenum tissue was provided by Tom Freeman of the Sanger
                    Center.
                    TAG_LIB=UI-R-EB1
                    TAG TISSUE=duodenum
                    TAG_SEQ=TGTGGTTCAT"
BASE COUNT       120 a      130 c      119 g      178 t
ORIGIN
```

US-10-045-180A-6 (1-31) x BQ200749 (1-547)

Alignment Scores:
Pred. No.:   2
Score: 0.557
Percent Similarity: 84.00   Length: 547
Best Local Similarity: 55.17%   Matches: 14
Query Match: 43.98%   Conservative: 2
                                              Mismatches: 13
                                              Indels: 0
                                              Gaps: 0

```
Qy    2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle  21
          ||  |  ||   |  ||     |||  | | | ||| || | | || |    |  | ||
Db  295 TGCTCTTGTAGAACCTCAAGCTGTCGTTTGGAGAAGGCTCTCTGGGGCATGTCGTCTC  236

Qy   22 LeuGlyGluArgTyrProIleCysCys  30
          ||| | ||  || |||   ||  |:::
Db  235 AATGGCGGCATCTACCGACTCTGTTGC  209
```

RESULT 6
LOCUS   AW467913   287 bp   mRNA   linear   EST 24-FEB-2000
DEFINITION he28d03.x1 NCI_CGAP_CML1 Homo sapiens cDNA clone IMAGE:2920325 3'
            similar to gb:M26602 NEUTROPHIL DEFENSINS 1, 2 AND 3 PRECURSOR
            (HUMAN);, mRNA sequence.
ACCESSION   AW467913
VERSION   AW467913.1 GI:7038019
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 287)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Elisabeth Paietta, Jonathan D. Licht, M.D.,
            Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
            Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq Primer: -40UP from Gibco
            High quality sequence stop: 231.
FEATURES   Location/Qualifiers
     source   1..287
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2920325"
                     /tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
                     rearrangement positive, includes both chronic phase and
                     myeloid blast crisis"
                     /lab_host="DH10B"

Right column

```
                    /clone_lib="NCI_CGAP_CML1"
                    /note="Organ: whole blood, Vector: pCMV-SPORT6; Site_1:
                    SalI; Site_2: NotI; Cloned unidirectionally. Primer:
                    Oligo dT. Library constructed by Life Technologies."
BASE COUNT        71 a       65 c       70 g       80 t        1 others
ORIGIN
```

US-10-045-180A-6 (1-31) x AW467913 (1-287)

Alignment Scores:
Pred. No.:   2
Score: 0.427
Percent Similarity: 82.00   Length: 287
Best Local Similarity: 55.17%   Matches: 13
Query Match: 42.93%   Conservative: 3
                                              Mismatches: 13
                                              Indels: 0
                                              Gaps: 0

```
Qy    2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle  21
          || |:: |   |  ||      |  | | ||| || | |  |   ||    ||  | |
Db  211 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGGAACGTTCNGCCATGGAAGGAACCTGC  152

Qy   22 LeuGlyLeuArgTyrProIleCysCys  30
          ||| | |  |  |||    ||  |:::
Db  151 CAGGGAAGACTCTGGGCATTCTGCTGC  125
```

RESULT 7
LOCUS   BI021835   349 bp   mRNA   linear   EST 14-JUN-2001
DEFINITION RC5-MT0259-170101-011-A09 MT0259 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI021835
VERSION   BI021835.1 GI:14428465
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 349)
  AUTHORS   Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
            Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
            Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
            Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
            M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
            Simpson, A.J.
  TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
            Ludwig Institute for Cancer Research
            Laboratory of Cancer Genetics
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC5&t2=RC5-MT0259-
            170101-011-A09&t3=20001-01-17&t4=1)
            Seq Primer: puc 18 forward
            High quality sequence stop: 274.
FEATURES   Location/Qualifiers
     source   1..349
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0259"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles

```
                              into the pUC 18 vector. Reverse transcription of tissue
                              mRNA and cDNA amplification were performed under low
                              stringency conditions."
BASE COUNT     88 a     91 c     96 g     73 t      1 others
ORIGIN

US-10-045-180A-6 (1-31) x BI021835 (1-349)

Alignment Scores:
  Pred. No.:              0.555
  Score:                 82.00
  Percent Similarity:    56.67%
  Best Local Similarity: 43.33%
  Query Match:           42.93%       Length:      349
  DB:                    12           Matches:      13
                                      Conservative:  4
                                      Mismatches:   13
                                      Indels:        0
                                      Gaps:          0

Qy    22 LeuGlyGluArgThrTyrProIleCysGlyTyr 31
         |||  ::|  |||
Db   257 CAGGGAAGACTCTGGGCATTCTGCTGCTGG 286

Qy     2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
         ||::|| :||::|||
Db   197 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAGAACGTCGCTATGGAACCTGCATCTAC 256

RESULT 8
LOCUS         BQ334373                    181 bp    mRNA       linear        EST 20-MAY-2002
DEFINITION    BQ334373/c QV1-MT0166-131100-482-a05 MT0166 Homo sapiens cDNA, mRNA sequence.
ACCESSION     BQ334373
VERSION       BQ334373.1   GI:20989461
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 181)
  AUTHORS     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
              Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
              ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.
  TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
  JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE     20202663
  PUBMED      10737800
COMMENT       Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2707001
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPESP/LICR Human Cancer Genome
              Project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/scripts/getthtml2.pl?ti1=QV1&t2=QV1-MT0166-
              131100-482-a05&t3=2000-11-13&t4=1)
              Seq primer: puc 18 forward
              High quality sequence stop: 7.
FEATURES             Location/Qualifiers
     source          1..181
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0166"
                     /note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
```

```
                              mRNA and cDNA amplification were performed under low
                              stringency conditions."
BASE COUNT     34 a     48 c     44 g     55 t
ORIGIN

US-10-045-180A-6 (1-31) x BQ334373 (1-181)

Alignment Scores:
  Pred. No.:              0.42
  Score:                 80.00
  Percent Similarity:    55.17%
  Best Local Similarity: 44.83%
  Query Match:           41.88%       Length:      181
  DB:                    13           Matches:      13
                                      Conservative:  3
                                      Mismatches:   13
                                      Indels:        0
                                      Gaps:          0

Qy    22 LeuGlyGluArgThrTyrProIleCysCys 30
         |||   :::  ||||
Db    90 CAGGGAAGACTCTGGGCATTCTGCTGC 64

Qy     2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
         ||::|| :||::|||
Db   150 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAGAACGTCGCTATGGAACCTGCATCTAC 91

RESULT 9
LOCUS         BE831361                    219 bp    mRNA       linear        EST 22-SEP-2000
DEFINITION    QV4-MT0050-220600-262-c11 MT0050 Homo sapiens cDNA, mRNA sequence.
ACCESSION     BE831361
VERSION       BE831361.1   GI:10263752
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 219)
  AUTHORS     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
              Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
              ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.
  TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
  JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE     20202663
  PUBMED      10737800
COMMENT       Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2707001
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPESP/LICR Human Cancer Genome
              Project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/scripts/getthtml2.pl?ti1=QV4-MT0050-220-
              600-262-c11&t3=2000-06-22&t4=1)
              Seq primer: puc 18 forward
              High quality sequence stop: 219.
FEATURES             Location/Qualifiers
     source          1..219
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0050"
                     /note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
```

BASE COUNT   54 a   58 c   62 g   45 t
ORIGIN

"stringency conditions."

US-10-045-180A-6 (1-31) x BB831361 (1-219)

Alignment Scores:
Pred. No.:
Score: 0.541   Length: 219
Percent Similarity: 80.00   Matches: 13
Best Local Similarity: 55.17%   Conservative: 3
Query Match: 44.83%   Mismatches: 13
DB: 41.88%   Indels: 0
                    10   Gaps: 0

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
           |::|||:::::||:   :::||||||||||
Db   125 TGCTATTGCAGAATACCAGCCGTGCAATTGCAGGAGAACGTGCTATGGAACCTGCATCTAC 184

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
           ||| :::||
Db   185 CAGGGGAAGACTCTGGGCCATTCTGCTGC 211

RESULT 10
LOCUS   AA321199   225 bp   mRNA   linear
DEFINITION   EST24002 Bone marrow Homo sapiens cDNA 5' end similar to defensin
               1, mRNA sequence.
ACCESSION   AA321199
VERSION   AA321199.1 GI:1973846
KEYWORDS   EST.
SOURCE   Homo sapiens
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 225)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
               ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
               ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
               Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
               ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
               Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
               Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
               Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
               Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
               Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F, Li,Y.,
               Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
               Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
               Hu,J.S., Greene,J.M, Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
               Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
               Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
               ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
               Venter,J.C.
  TITLE   Initial assessment of human gene diversity and expression patterns
               based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
  PUBMED   7566098
  COMMENT   Other_ESTs: THC169174
               Contact: Kerlavage,AR
               The Institute for Genomic Research
               Bioinformatics
               9712 Medical Center Drive, Rockville, MD 20850 USA
               Tel: 301869056
               Fax: 301869423
               Email: arkerlav@tigr.org
               For clone availability, additional sequence and expression
               information related to this EST, please check the TIGR Human Gene
               Index (http://www.tigr.org/tdb/hgi/hgi.html)
               Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
     source   1..225
                     /organism="Homo sapiens"
                     /mol_type="mRNA"

BASE COUNT   65 a   52 c   56 g   52 t
ORIGIN

                     /db_xref="ATCC (inhost):122070"
                     /db_xref="taxon:9606"
                     /sex="mixed"
                     /tissue_type="bone marrow"
                     /dev_stage="adult"
                     /clone_lib="Bone marrow"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"

US-10-045-180A-6 (1-31) x AA321199 (1-225)

Alignment Scores:
Pred. No.:
Score: 0.561   Length: 225
Percent Similarity: 80.00   Matches: 13
Best Local Similarity: 55.17%   Conservative: 3
Query Match: 44.83%   Mismatches: 13
DB: 41.88%   Indels: 0
                    9   Gaps: 0

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
           |::|||:::::||:   :::||||||||||
Db   64 TGCTATTGCAGAATACCAGCCGTGCAATTGCAGGAGAACGTGCTATGGAACCTGCATCTAC 123

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
           ||| :::||
Db   124 CAGGGGAAGACTCTGGGCCATTCTGCTGC 150

RESULT 11
LOCUS   AA321481   229 bp   mRNA   linear
DEFINITION   EST239961 Bone marrow Homo sapiens cDNA 5' end similar to defensin
               1, mRNA sequence.
ACCESSION   AA321481
VERSION   AA321481.1 GI:1973808
KEYWORDS   EST.
SOURCE   Homo sapiens
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 229)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
               ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
               ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
               Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
               ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
               Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
               Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
               Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
               Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
               Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F, Li,Y.,
               Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
               Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
               Hu,J.S., Greene,J.M, Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
               Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
               Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
               ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
               Venter,J.C.
  TITLE   Initial assessment of human gene diversity and expression patterns
               based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
  PUBMED   7566098
  COMMENT   Other_ESTs: THC169174
               Contact: Kerlavage,AR
               The Institute for Genomic Research
               Bioinformatics
               9712 Medical Center Drive, Rockville, MD 20850 USA
               Tel: 301869056
               Fax: 301869423
               Email: arkerlav@tigr.org
               For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

FEATURES   Location/Qualifiers
     source   1..229
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):122032"
                     /db_xref="taxon:9606"
                     /sex="mixed"
                     /tissue_type="bone marrow"
                     /dev_stage="adult"
                     /clone_lib="Bone marrow"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT   66 a   54 c   49 g   60 t
ORIGIN

US-10-045-180A-6 (1-31) x AA321481 (1-229)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
        |::|||||||||:::|
Db   22 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTATGGAACCTGCATCTAC 81

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
        ||:||
Db   82 CAGGGAGAAGACTCTGGGCATTCTGCTGC 108

Alignment Scores:
Pred. No.:
Score: 0.575
Percent Similarity: 55.17%   Length: 229
Percent Identity: 44.83%   Matches: 13
                                         Conservative: 3
Best Local Similarity: Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: 9   Gaps: 0

RESULT 12
LOCUS   AW075767/c
DEFINITION xa85b06.x1 NCI_CGAP_CML1 Homo sapiens cDNA clone IMAGE:2573555 3'
            similar to gb:M26602 NEUTROPHIL DEFENSINS 1, 2 AND 3 PRECURSOR
            (HUMAN);, mRNA sequence.
ACCESSION   AW075767
VERSION   AW075767.1 GI:6030765
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 233)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Elisabeth Paietta, Jonathan D. Licht, M.D.;
            Michael R. Emmett-Buck, M.D., Ph.D. cDNA Library Preparation: Life
            Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 226.

LOCUS   xa85b06.x1   233 bp   mRNA   linear   EST 13-OCT-1999

FEATURES   Location/Qualifiers
     source   1..233
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2573555"
                     /tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL

rearrangement positive, includes both chronic phase and
myeloid blast crisis"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_CML1"
                     /note="Organ: whole blood; Vector: pCMV-SPORT6; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally. Primer:
                     Oligo dT. Library constructed by Life Technologies."

BASE COUNT   62 a   49 c   55 g   67 t
ORIGIN

US-10-045-180a-6 (1-31) x AW075767 (1-233)

Alignment Scores:
Pred. No.:
Score: 0.588
Percent Similarity: 55.17%   Length: 233
Percent Identity: 44.83%   Matches: 13
                                         Conservative: 3
Best Local Similarity: Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: 9   Gaps: 0

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
        |::|||||||||:::|
Db   22 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTATGGAACCTGCATCTAC 81

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
        ||:||
Db 147 CAGGGAGAAGACTCTGGGCATTCTGCTGC 121

RESULT 13
LOCUS   BF893374/c
DEFINITION QV3-MT0129-111100-427-a02 MT0129 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF893374
VERSION   BF893374.1 GI:12284833
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 234)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
            M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
            Ludwig Institute for Cancer Research
            Laboratory of Cancer Genetics
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/getmt.l2.pl?ti=QV3-MT0129-
            111100-427-a02&t3=2000-11-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 234.

LOCUS   BF893374   234 bp   mRNA   linear   EST 18-JAN-2001

FEATURES   Location/Qualifiers
     source   1..234
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0129"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:

RESULT 14

US-10-045-180A-6 (1-31) x BF893374 (1-234)

```
ORIGIN
BASE COUNT      65 a     48 c     55 g     66 t
```

Smal; A mini-library was made by cloning products derived
from ORSTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

```
Alignment Scores:
Pred. No.:
Score:                    0.592
Percent Similarity:       80.00        Matches:        234
Best Local Similarity:    55.17%       Conservative:     3
Query Match:              44.83%       Mismatches:      13
DB:                       41.88%       Indels:           0
                             10        Gaps:             0

Qy    2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyThrCysPheIle 21
        |||:|:||||||||||:::|||||:|||||||||||||||
Db  222 TGCTATTGCAGAATACCAGGCGTGCATTGCAGGAGAACGTCGCTATGGAACCTGCCATCTAC 163

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
        ||:|:|||||||||
Db  162 CAGGGAAGACTCTGGGCCATTCTGCTGC 136
```

```
LOCUS       BI028909            241 bp    mRNA    linear    EST 14-JUN-2001
BI028909/c
DEFINITION  PM3-MT0205-050301-004-d10 MT0205 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI028909
VERSION     BI028909.1  GI:14435539
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 241)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM3-MT0205-
            050301-004-d10&t3=2001-03-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 23
            High quality sequence stop: 241.
FEATURES             Location/Qualifiers
     source          1..241
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0205"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
```

RESULT 15

US-10-045-180A-6 (1-31) x BI028909 (1-241)

```
ORIGIN
BASE COUNT      61 a     55 c     62 g     63 t
```

Smal; A mini-library was made by cloning products derived
from ORSTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

```
Alignment Scores:
Pred. No.:
Score:                    0.615
Percent Similarity:       80.00        Matches:        241
Best Local Similarity:    55.17%       Conservative:     3
Query Match:              44.83%       Mismatches:      13
DB:                       41.88%       Indels:           0
                             12        Gaps:             0

Qy    2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyThrCysPheIle 21
        |||:|:||||||||||:::|||||:||||| ||||||||||||
Db  216 TGCTATTGCAGAATACCAGGCGTGCATTGCAGGAGAACGTCGCTATGGAACCTGCCATCTAC 157

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
        ::|:|||||||||
Db  156 CAGGGAAGACTCTGGGCCATTCTGCTGC 130
```

```
LOCUS       BF903426            246 bp    mRNA    linear    EST 18-JAN-2001
BF903426
DEFINITION  IL2-MT0179-181200-278-H04 MT0179 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF903426
VERSION     BF903426.1  GI:12294989
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 246)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL2-MT0179-
            181200-278-H04&t3=2000-12-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 54
            High quality sequence stop: 246.
FEATURES             Location/Qualifiers
     source          1..246
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0179"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
```

```
                Smal; A mini-library was made by cloning products derived
                from ORBSTES PCR (U.S. Letters Patent Application No. 196
                ,716 - Ludwig Institute for Cancer Research) profiles
                into the pUC 18 vector. Reverse transcription of tissue
                mRNA and cDNA amplification were performed under low
                stringency conditions."

BASE COUNT   69 a      55 c      57 g      65 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                  0.633          Length:          246
Percent Similarity:     80.00          Matches:          13
Best Local Similarity:  55.17%         Conservative:      3
Query Match:            44.83%         Mismatches:       13
DB:                     41.88%         Indels:            0
                        10             Gaps:              0

US-10-045-180A-6 (1-31) x BP903426 (1-246)

Qy      2  CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle  21
           |||::||||||||:::   |||||||         ||||||||
Db     66  TGCTRATTGCAGAATACCAGCGTGCATTGCAGGAGGAACGTCGCTANGGAACCTGCATCTAC  125

Qy     22  LeuGlyGluArgTyrProIleCysCys  30
           |||         |:::  ||||||
Db    126  CAGGGAAGACTCTGGGCATTCTGCTGC  152
```

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 14:50:33 ; Search time 116.25 Seconds

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

(without alignments)
719.850 Million cell updates/sec

Title:
Perfect Score: 191
Sequence: US-10-045-180A-6
                   1 ICHCRVLYCIFGEHLGGTCFILGSRYPICCY 31

Scoring table: BLOSUM62
                   Xgapop 10.0   Xgapext 0.5
                   Ygapop 10.0   Ygapext 0.5
                   Fgapop 6.0   Fgapext 7.0
                   Delop   6.0   Delext   7.0

Searched: 2552756 seqs, 134971901 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                   Maximum Match 100%
                   Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model
-Q=/cgn2_1/USPTO_spool_p/US10045180/runat 17122003 145023 12520/app_query.fasta_1.860
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX-rng -MINMATCH=0.1 -LOOPCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10045180 @CGN_1_1_835 @runat 17122003 145023_12520 -ICPU=3
-NO_MMAP -LARGEQUERY _NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
        N_Geneseq_19Jun03:*
  1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
  2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
  3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
  4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
  5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
  6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
  7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
  8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
  9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
 10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
 11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
 12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
 13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
 14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
 15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
 16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
 17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
 18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
 19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
 20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
 21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
 22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
 23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
 24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
 25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	191	100.0	453	20	AAX26697	cDNA sequence enco
2	191	100.0	453	20	AAX26698	cDNA sequence enco
3	191	100.0	4415	20	AAX26696	Nucleotide sequenc
4	80	41.9	90	20	AAV83787	Antimicrobial pept
5	80	41.9	108	17	AAT17899	DNA encoding amphi
6	80	41.9	110	21	AAZ29363	Antimicrobial pept
7	80	41.9	110	24	ABL60444	DNA fragment of th
8	80	41.9	159	22	AAI88805	DNA polynucleoti
9	80	41.9	464	21	AAF20911	Human defensin 3 p
10	80	41.9	464	21	AAA34789	Human adenosine re
11	80	41.9	478	21	AAC04468	Human secreted pro
12	80	41.9	498	12	AAQ14908	Chronic myelogenou
13	80	41.9	498	24	AAB06601	Human defensin 1 c
14	80	41.9	509	21	AAC00098	Human secreted pro
15	80	41.9	514	24	ABN59722	Novel human coding
16	80	41.9	543	25	ABZ36767	Human GENSET codin
17	80	41.9	644	21	AAF21446	Human defensin 3 p
18	80	41.9	652	24	ABZ11202	Human polynucleoti
19	80	41.9	664	21	ABZ11183	Human polynucleoti
20	80	41.9	1097	15	AAQ68911	Human c-myc far up
21	80	41.9	3710	21	AAF20910	Human defensin 3 p
22	80	41.9	3710	21	AAA34788	Human adenosine re
23	80	41.9	3710	24	ABQ80602	Human defensin 1-r
24	80	41.9	3710	24	ABK83829	Human cDNA differe
25	80	41.9	8365	21	AAF21445	Human defensin 2 p
26	80	41.9	209273	21	AAF21437	Human factor-relat
27	74	38.7	125	17	AAT17900	DNA encoding amphi
28	74	38.7	411	24	ABQ60009	Human colon cancer
29	74	38.7	412	24	ABK39009	cDNA encoding lung
30	74	38.7	412	25	ACA11338	Human lung adenoca
31	74	38.7	412	24	AAA34788	Human lung adenoca
32	74	38.7	424	14	ACA02524	Sequence encoding
33	74	38.7	449	24	AAQ53216	Sequence encoding
34	74	38.7	514	22	ABQ60783	Human defensin 5 D
35	74	38.7	514	22	AAH57427	Human defensin D
36	74	38.7	586	24	ABQ60212	Human intestine ce
37	74	38.7	967	20	AAX08992	Human colon cancer
38	74	38.7	2880	14	AAQ53218	Dual selection cas
39	71	37.2	102	14	AAQ50106	Genomic sequence
40	71	37.2	102	14	AAQ50111	Genomic sequence
41	71	37.2	156	14	AAE16106	Corticostatin. Or
42	69.5	36.4	299	16	AAT30742	Corticostatin. Sy
43	69	36.1	365	17	AAT30745	Human gene signatu
44	69	36.1	420	17	AAT30744	Mouse cryptdin 4 c
45	67	35.1	342	17	AAT30746	Mouse cryptdin 7 c
44	67	35.1	388	17	AAT30751	Mouse cryptdin 8 c
45	66.5	34.8	452	14	AAQ53217	Mouse cryptdin 13
						Sequence encoding

ALIGNMENTS

RESULT 1
AAX26697
ID   AAX26697 standard; cDNA; 453 BP.
XX
AC   AAX26697;
XX
DT   18-JUN-1999 (first entry)
XX
DE   cDNA sequence encoding human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis; ss.
XX
OS   Homo sapiens.
XX

PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997; 97FR-0010823.
XX
PR   29-AUG-1997; 97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
DR   P-PSDB; AAY01604.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Disclosure; Fig 3; 56pp; French.
XX
CC   The present sequence encodes human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The Def-X polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can be used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 453 BP; 107 A; 121 C; 97 G; 128 T; 0 other;

Alignment Scores:
Pred. No.:   2.99e-17
Score: 191.00   Length: 453
Percent Similarity: 100.00%   Matches: 31
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 20   Mismatches: 0
DB: Indels: 0
                                          Gaps: 0

US-10-045-180A-6 (1-31) x AAX26697 (1-453)

Qy   1 IleCysHisCysArgValLeuTyrCysIlePheGlyGlyHisLeuGlyGlyThrCysPhe 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 ATCTGCCATTGCAGAGTACTATACTGCATTTTGGAGAACATCTTGGTGGGACCTGCTTC 300

Qy 21 IleLeuGlyGluArgTyrProIleCysCysTyr 31
         |||||||||||||||||||||||||||||||||
Db 301 ATCCTTGGTGAACGCTACCCAATCTGCTGCTAC 333

RESULT 2
AAX26698
ID   AAX26698 standard; cDNA; 453 BP.
XX
AC   AAX26698;
XX
DT   18-JUN-1999 (first entry)
XX
DE   cDNA sequence encoding human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair; cancer; melanoma;
KW   endocrine regulation; corticostatic regulation; cancer;
KW   AIDS; immune deficiency; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997; 97FR-0010823.
XX
PR   29-AUG-1997; 97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Claim 5; Fig 4; 56pp; French.
XX
CC   The present sequence encodes human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The Def-X polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can be used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 453 BP; 107 A; 121 C; 97 G; 128 T; 0 other;

Alignment Scores:
Pred. No.:   2.99e-17
Score: 191.00   Length: 453
Percent Similarity: 100.00%   Matches: 31
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 20   Mismatches: 0
DB: Indels: 0
                                          Gaps: 0

US-10-045-180A-6 (1-31) x AAX26698 (1-453)

Qy   1 IleCysHisCysArgValLeuTyrCysIlePheGlyGlyHisLeuGlyGlyThrCysPhe 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 ATCTGCCATTGCAGAGTACTATACTGCATTTTGGAGAACATCTTGGTGGGACCTGCTTC 300

Qy 21 IleLeuGlyGluArgTyrProIleCysCysTyr 31
         |||||||||||||||||||||||||||||||||
Db 301 ATCCTTGGTGAACGCTACCCAATCTGCTGCTAC 333

RESULT 3
AAX26696
ID   AAX26696 standard; DNA; 4415 BP.
XX
AC   AAX26696;
XX
DT   18-JUN-1999 (first entry)
XX
DE   Nucleotide sequence of human defensin (Def-X).
XX

KW Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW cytostatic; anticancer; inflammation; tissue repair;
KW endocrine regulation; corticostatic regulation; tissue repair;
KW AIDS; immune deficiency; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN FR276783.2-A1.
XX
PD 05-MAR-1999.
XX
PF 29-AUG-1997;   97FR-0010823.
XX
PR 29-AUG-1997; 97FR-0010823.
XX
PA (GBST) GENSET SA.
XX
PI Bougueleret L, Chumakov I;
XX
DR WPI; 1999-183266/16.
XX
PS Claim 5; Fig 2; 56pp; French.
XX
PT Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT anticancer agent, pesticide, etc.
XX
CC The present sequence encodes human defensin (Def-X). The Def-X
CC polypeptide can be used as an antimicrobial, antiparasitic agent or
CC a pesticide. The Def-X polypeptide can be used as a cytostatic agent
CC (especially anticancer) agent, and as an agent for modulating processes
CC of inflammation. The polypeptide can be used in a composition for external
CC regulation. The polypeptide can be used in a composition for external
CC topical use, especially in a cosmetic composition. Compositions
CC containing Def-X can be used for prevention and treatment of microbial
CC and parasitic infections, especially where the microbial or parasitic
CC infections are Gram-positive or -negative bacterial infections or
CC mycobacterial, fungal or spirochaete infections, or where the viral
CC infections are associated with enveloped viruses, especially HSV and HIV.
CC The compositions can used for prevention and/or treatment of cancers,
CC especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC lung cancer or colorectal carcinoma, and for enhancing immunity,
CC especially in the case of AIDS, or preventing immune deficiency,
CC especially in the treatment of psoriasis, as well as for modulating
CC inflammatory processes, especially in the case of chronic inflammatory
CC disorders.
XX
SQ Sequence 4415 BP; 1128 A; 1109 C; 912 G; 1252 T; 14 other;

Alignment Scores:
Pred. No.:
Score: 4.63e-16
Pred. No.:   191.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 20

US-10-045-180A-6 (1-31) x AAX26696 (1-4415)

Qy   1 IleCysHisCysArgValLeuTyrCysIlePheGlyGlyHisLeuGlyGlyThrCysPhe 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4181 ATCTGCCATTGCAGAGTACTATACTGCATTTTGGAGAACATCTTGGTGGGACCTGCTTC 4240

Qy   21 IleLeuGlyGluArgTyrProIleCysCysTyr 31
              |||||||||||||||||||||||||||||||||
Db   4241 ATCCTTGGTGAACGCTACCAAATCTGCTGCTAC 4273

XX
DE Antimicrobial peptide HNP-I encoding DNA.
XX
KW Antimicrobial; fusion; acidic peptide; recombinant; microorganism;
KW guamerin; basic peptide; HNP-I; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   1..90
FT   /*tag= a
FT   /note= "the start and stop codons are not indicated"
XX
PN WO9854336-A1.
XX
PD 03-DEC-1998.
XX
PF 28-MAY-1998; 98WO-KR00132.
XX
PR 09-APR-1998; 98KR-0013372.
PR 28-MAY-1997; 97KR-0021312.
XX
PA (KOAD) KOREA ADV INST SCI & TECHNOLOGY.
PA (SAMY-) SAMYANG GENEX CORP.
XX
PI Hong S, Kang MH, Kim JH, Kim S, Lee H, Lee JH;
XX
DR WPI; 1999-059844/05.
DR P-PSDB; AAW87608.
XX
PS Example 6; Page 18; 52pp; English.
XX
PT New method for mass production of antimicrobial peptides - by
PT constructing fusion genes comprising acidic and antimicrobial
PT peptide genes and transforming host with vector containing these
CC The invention relates to mass production of antimicrobial peptides. The
CC method comprises constructing a fusion gene containing a first gene
CC encoding a negatively charged acidic peptide having at least two cysteine
CC residues, and a second gene encoding a positively charged basic
CC antimicrobial peptide. A host microorganism is transformed with a vector
CC containing the fusion gene and then cultured. The expressed antimicrobial
CC peptide is then recovered. The method is used to mass produce
CC antimicrobial peptides in recombinant microorganisms. The inhibitory
CC effect of the expressed antimicrobial peptide upon the growth of the host
CC microorganism is considerably reduced by fusing it to the acidic peptide.
CC Therefore, the use of the fusion gene provides an economic, recombinant
CC alternative of mass producing antimicrobial peptides, which overcomes the
CC disadvantages of low-productivity and poor economy, previously
CC encountered by recombinant and chemical methods. The present sequence
CC represents the DNA encoding an antimicrobial peptide HNP-I. This
CC can be used along with the acidic peptide Guamerin gene in the
CC construction of the fusion gene.
XX
SQ Sequence 90 BP; 9 A; 27 C; 31 G; 23 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 0.00784
Pred. No.:   80.00
Percent Similarity: 55.17%
Best Local Similarity: 44.83%
Query Match: 20

US-10-045-180A-6 (1-31) x AAV83787 (1-90)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGlyHisLeuGlyGlyThrCysPheIle 21
              || ||:||||:::||||:|||| |:||||||||||||||||||||:||||:::||||
Db   4 TGCTACTGCCGTATCCCGCGCTGCATGCGCGGTGAGCGGTCGTTACGGTACCTGCATCTAC 63

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
              ||| ||:
Db   LeuGlyGluArgTyrProIleCysCys

Db

RESULT 5

AAT17899

64   CAGGGTTCGTCGTGTGGGGTTCGTCTGC 90

ID   AAT17899 standard; DNA; 108 BP.
XX
AC   AAT17899;
XX
DT   24-SEP-1996 (first entry)
XX
DE   DNA encoding amphiphilic peptide human NP-1.
XX
KW   Human NP-1; amphiphilic peptide; recombinant production;
KW   protease deficient; microbial host cell; expression vector;
KW   Escherichia coli; K-12 cell; vector; cleavable fusion protein;
KW   carbohydrate binding protein; anti-parasitic; anti-fungal;
KW   anti-tumour; anti-viral; anti-cancer; anti-microbial; ss.
XX
OS   Synthetic.
XX
FH
FH   Key   Location/Qualifiers
FT
FT   CDS   6..101
FT   /tag= a
XX
PN   WO9604373-A2.
XX
PD   15-FEB-1996.
XX
PF   26-JUL-1995; 95WO-US10219.
XX
PR   29-JUL-1994; 94US-0282030.
XX
PA   (MAGA-) MAGAININ PHARM INC.
XX
PI   Anderson GM, Kari P, Pierce JC, Williams JI;
XX
DR   WPI; 1996-129390/13.
DR   P-PSDB; AAR92824.
XX
PT   Recombinant production of amphiphilic peptide in protease deficient
PT   microbial host, pref. E.coli K-12 - useful in prodn. of
PT   antimicrobial, antiviral and anticancer peptide(s)
XX
PS   Disclosure; Page 24; 103pp; English.
XX
CC   The present sequence, which encodes human NP-1 (an amphiphilic
CC   peptide) is used in 2 novel methods for the recombinant prodn. of
CC   human NP-1. The 1st method comprises transforming a protease
CC   deficient (PD) microbial host cell with an expression vector contg.
CC   the DNA, under the control of a regulatory sequence operable in the
CC   host, and expressing the peptide in the transformed host. The 2nd
CC   method comprises transforming an E. coli PD K-12 cell with a vector
CC   that expresses a cleavable fusion protein, comprising at least part
CC   of a carbohydrate binding protein (CBP) and the peptide, expressing
CC   the fusion protein in the cell and cleaving the protein to obtain
CC   the peptide substantially free of CBP residues. These methods for
CC   producing and processing human NP-1 allow high levels of the
CC   peptide to accumulate in certain PD microbial host cells, despite
CC   the peptides anti-microbial potency, and efficient recovery of the
CC   full length peptide. The peptide produced, unlike most natural
CC   analogous peptides, exhibits a broader range of activity and/or
CC   greater potency compared to a related natural peptide, e.g. the
CC   peptide may be used as an anti-parasitic, anti-fungal, anti-tumour,
CC   anti-cancer or an anti-viral agent.
XX
SQ   Sequence 108 BP; 20 A; 25 C; 28 G; 35 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 0.00977
Score: 80.00   Length: 108
Percent Similarity: 55.17%   Matches: 13
Best Local Similarity: 44.83%   Conservative: 3
Query Match: 41.88%   Mismatches: 13
                                               Indels: 0

DB: 17   Gaps: 0

US-10-045-180a-6 (1-31) x AAT17899 (1-108)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyLeuHisLeuGlyThrCysProIle 21
        |||::|||||||:|||||||||||||||||||||||||
Db 12 TGTTACTGCCGTATTCCGGGCATGCAATCGCAGGCGAGGCGTCGCTATGGTACTTGTATTAC 71

Qy 22 LeuGlyGluArgTyrProIleCysCys 30
        |||| ::: |||||
Db 72 CAGGGTTCGTCGTGTGGGGCATTCGTGTGC 98

Db

RESULT 6

AAA29363

ID   AAA29363 standard; DNA; 110 BP.
XX
AC   AAA29363;
XX
DT   29-FEB-2000 (first entry)
XX
DE   Antimicrobial peptide, human alpha defensin encoding DNA.
XX
KW   purF gene; glutamine pyrophosphoribosyl pyrophosphate amidotransferase;
KW   purF derivative; fusion partner; antimicrobial peptide; alpha defensin;
KW   mass production; cleavage site; hydroxylamine; CNBr; DNA construct;
KW   neutralise; toxicity; pharmaceutical industry; food industry; ds.
XX
OS   Homo sapiens.
XX
FH
FH   Key   Location/Qualifiers
FT   CDS   11..103
FT   /tag= a
FT   /product= "Human alpha defensin peptide"
FT   /note= "Antimicrobial peptide used in DNA construct"
XX
PN   WO9964611-A1.
XX
PD   16-DEC-1999.
XX
PF   08-JUN-1999; 99WO-KR00282.
XX
PR   09-JUN-1998; 98KR-0022117.
PR   14-MAY-1999; 99KR-0017920.
XX
PA   (SAMY-) SAMYANG GENEX CORP.
XX
PI   Kim JH, Kang MH, Lee J, Park SH, Lee JW, Hong SS, Lee H;
XX
DR   WPI; 2000-097542/08.
DR   P-PSDB; AAY44323.
XX
PT   New DNA constructs useful for mass production of antimicrobial peptides
PT   in microorganism hosts -
XX
PS   Claim 1; Fig 1; 67pp; English.
XX
CC   The present DNA sequence encodes an antimicrobial peptide, human alpha
CC   defensin. It is used along with a derivative of purF gene sequence that
CC   functions as a fusion partner. A DNA construct that comprises, this
CC   antimicrobial peptide encoding sequence and the entire, partial or
CC   derivative of purF gene, is used for mass production of the antimicrobial
CC   peptide in microorganisms without killing the host cells. Use of the purF
CC   gene derivative sequence, neutralises the toxicity of the antimicrobial
CC   peptides against the host microorganism. The antimicrobial peptides are
CC   useful commercially in the pharmaceutical and food industries.
XX
SQ   Sequence 110 BP; 14 A; 31 C; 37 G; 28 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 0.00998
Score: 80.00   Length: 110
Percent Similarity: 55.17%   Matches: 13
Percent Similarity: 44.83%   Conservative: 3
Best Local Similarity: 44.83%   Mismatches: 13

Left column

Query Match: Indels: 0
DB: 41.88%   21   Gaps: 0

US-10-045-180A-6 (1-31) x AA229363 (1-110)

```
Qy   2  CysHisCysArgValLeuTyrCysIlePheGlyGluHisIleuGlyGlyThrCysPheIle  21
           |||  ||| |:::||:|||  ||| :|:|
Db  14  TGCTACTGCCGTANTCCCGGCCGTGCATCGCGGGGTGAGCGCGTCGTTACGGTGTACCTGCATCTAC  73

Qy  22  LeuGlyGluArgTyrProIleCysCys  30
           ||| ||| |||
Db  74  CAGGGTCGTCTGTGTGGGCGTTCTGCTGC  100
```

RESULT 7

ID   ABL60444
AC   ABL60444;

ABL60444 standard; DNA; 110 BP.

DT   28-MAR-2003 (first entry)

DE   DNA fragment of the invention HNP-I.

KW   Gene expression; peptide antibiotic; purF gene; ds.

OS   Unidentified.

FH   Key   Location/Qualifiers
FT   CDS   11..103
FT   /*tag= a
FT   /partial
FT   /note="no start codon present"

PN   KR200109B973-A.

PD   08-NOV-2001.

PF   08-JUN-2001; 2001KR-0031889.
PR   08-JUN-2001; 2001KR-0031889.

PA   (SAMY-) SAMYANG GENEX CORP.

PI   Hong SS, Kang MH, Kim JH, Lee HS, Lee JH, Lee JW, Park SH;

DR   WPI; 2002-301977/34.
DR   P-PSDB; ABB81939.

PT   Gene expression system useful for mass-production of peptide
PT   antibiotics and vectors derived from microorganisms -

PS   Disclosure; Page 16; 56pp; Korean.

CC   The invention relates to a genetic component which mass-produces peptide
CC   antibiotics effectively from microorganisms. The genetic component
CC   consists of a first gene sequence which codes for the whole or partial
CC   purF gene or its derivative, and a second gene sequence which codes
CC   peptide antibiotics. The mass-production method of peptide antibiotics
CC   comprises the steps of; constructing an expression vector including the
CC   genetic component, transforming a bacterial host cell with the vector,
CC   culturing the transformed cell to express the genetic component, and
CC   recovering the peptide antibiotics. The expression vector is selected
CC   from the group consisting of pGNX2, pGNX3, pGNX4 and pGNX5, and it has
CC   a high copy number of origin, strong transcription promoter and
CC   structural gene. The sequences given in records ABL60400-ABL60464
CC   represent DNA sequences of the invention.

SQ   Sequence 110 BP; 14 A; 31 C; 37 G; 28 T; 0 other;

Alignment Scores:
Pred. No.:   0.00998   Length: 110
Score: 80.00   Matches: 13
Percent Similarity: 55.17%   Conservative: 3

Right column

Best Local Similarity: 44.83%
Query Match: 41.88%
DB: 24

US-10-045-180A-6 (1-31) x ABL60444 (1-110)

Mismatches: 13
Indels: 0
Gaps: 0

```
Qy   2  CysHisCysArgValLeuTyrCysIlePheGlyGluHisIleuGlyGlyThrCysPheIle  21
           |||  ||| |:::||:|||  ||| :|:|
Db  14  TGCTACTGCCGTANTCCCGGCCGTGCATCGCGGGGTGAGCGCGTCGTTACGGTGTACCTGCATCTAC  73

Qy  22  LeuGlyGluArgTyrProIleCysCys  30
           ||| ||| |||
Db  74  CAGGGTCGTCTGTGTGGGCGTTCTGCTGC  100
```

RESULT 8

ID   AA188805
AC   AA188805;

AA188805 standard; cDNA; 159 BP.

DT   06-NOV-2001 (first entry)

DE   Human polynucleotide SEQ ID NO 8865.

KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; 88.

OS   Homo sapiens.

PN   WO200164835-A2.

PD   07-SEP-2001.

PF   26-FEB-2001; 2001WO-US04927.
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.

PA   (HYSE-) HYSEQ INC.

PI   Tang YT, Liu C, Drmanac RT;

DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO08874.

PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -

PS   Claim 1; SEQ ID NO 8865; 1399pp + Sequence Listing; English.

CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy. The
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.

SQ   Sequence 159 BP; 49 A; 34 C; 44 G; 32 T; 0 other;

Alignment Scores:
Pred. No.:   0.0156   Length: 159
Score: 80.00   Matches: 13

```
                              Percent Similarity:  55.17%     Conservative:   3
                              Best Local Similarity:  44.83%  Mismatches:    13
                              Query Match:          41.88%    Indels:         0
                              DB:                              Gaps:          0

US-10-045-180A-6 (1-31) x AAI88805 (1-159)

Qy   2 CysHisCysArgValLeuThrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
         |::||:||:|::|
Db  15 TGCTATTGCAGAATACCAGGCGTGCATTGCACGAGGAACGTCGCTATGGAACCTGCATCTAC 74

Qy  22 LeuGlyGluArgTyrProIleCysCys 30
          |||||
Db  75 CAGGGAAGAGACTTCTGGGCATTCTGCTGC 101
```

RESULT 9
ID   AAF20911 standard; DNA; 464 BP.
XX
AC   AAF20911;
XX
DT   14-MAR-2001 (first entry)
XX
DE   Human defensin 3 polynucleotide fragment #2478.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999; 99US-0127958.
XX
PI   Nyce JW.
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 218; 1592pp; English.

The present invention describes low adenosine (A) content antisense
oligonucleotides and compositions (I) comprising them. In the antisense
oligonucleotides the A is replaced by a 'universal' or alternative base.
(I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
The antisense oligonucleotides and (I) can be used to down-regulate the
expression and or activity of target polypeptides associated with
lung/respiratory disorders and malignancies, such as stimulating and
activating peptide factors and transmitters, transcription factors,
immunoglobulins and antibodies, antibody receptors, cytokines and
chemokines, endogenously produced specific and non-specific enzymes,
binding proteins, adhesion molecules and their receptors, cytokine and
chemokine receptors, adenosine receptors, bradykinin receptors, central
nervous system (CNS) and peripheral nervous and non-nervous system

CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 464 BP; 110 A; 128 C; 119 G; 107 T; 0 other;

```
Alignment Scores:
Pred. No.:
Score:                 0.0564
                      80.00
Percent Similarity:   55.17%      Length:        464
Best Local Similarity: 44.83%     Matches:        13
Query Match:          41.88%      Conservative:    3
DB:                     21         Mismatches:     13
                                   Indels:          0
                                   Gaps:            0
```

US-10-045-180A-6 (1-31) x AAF20911 (1-464)

```
Qy   2 CysHisCysArgValLeuThrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
         |::||:||:|::|
Db 285 TGCTATTGCAGAATACCAGGCGTGCATTGCACGAGGAGAACGTCGCTATGGAACCTGCATCTAC 344

Qy  22 LeuGlyGluArgTyrProIleCysCys 30
          |||||
Db 345 CAGGGAAGAGACTTCTGGGCATTCTGCTGC 371
```

RESULT 10
ID   AAA34789 standard; DNA; 464 BP.
XX
AC   AAA34789;
XX
DT   28-JUL-2000 (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2478.
XX
KW   Low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO20009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999; 99WO-US17712.
XX
PR   03-AUG-1998; 98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,

PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 612; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 464 BP; 110 A; 128 C; 119 G; 107 T; 0 other;

US-10-045-180A-6 (1-31) x AAA34789 (1-464)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
        ||:::||||||||||||||||:::
Db   285 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGGAGAACGTGCGCTATGGAACCTGCATCTAC 344

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
        |||:::
Db   345 CAGGGGAAGACTCTGGGGCATTCTGCTGC 371

Alignment Scores:
Pred. No.:
Score: 0.0564
Percent Similarity: 80.00
Best Local Similarity: 55.17%
Query Match: 41.88%   21

Length: 464
Matches: 13
Conservative: 3
Mismatches: 13
Indels: 0
Gaps: 0

RESULT 11
AAC04468
ID   AAC04468 standard; cDNA; 478 BP.
XX
AC   AAC04468;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 8543.
XX
OS   Homo sapiens.
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999; 99US-0122487.

XX
PA   (GEST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 8543; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 478 BP; 117 A; 129 C; 127 G; 101 T; 4 other;

US-10-045-180A-6 (1-31) x AAC04468 (1-478)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
        ||:::||||||||||||||||:::
Db   346 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGGAGAACGTGCGCTATGGAACCTGCATCTAC 405

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
        |||:::
Db   406 CAGGGGAAGACTCTGGGGCATTCTGCTGC 432

Alignment Scores:
Pred. No.:
Score: 0.0585
Percent Similarity: 80.00
Best Local Similarity: 55.17%
Query Match: 41.88%   21

Length: 478
Matches: 13
Conservative: 3
Mismatches: 13
Indels: 0
Gaps: 0

RESULT 12
AAQ14908
ID   AAQ14908 standard; cDNA; 498 BP.
XX
AC   AAQ14908;
XX
DT   25-FEB-1992 (first entry)
XX
DE   Chronic myelogenous leukaemia-derived mrs cDNA.
XX
KW   CML; Philadelphia chromosome; myeloid-related sequence;
KW   chromosome 8; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   16..378
FT   /*tag= a
FT   /note= "see comments"
FT   polyA_signal   479..484
FT   /*tag= b
XX
PN   US5067792-A.
XX
PD   19-NOV-1991.

XX
PF 02-DEC-1988; 88US-0279451.
XX
PR 02-DEC-1988; 88US-0279451.
PR 28-SEP-1984; 84US-0655942.
PR 04-SEP-1987; 87US-0094099.
XX
PA (TEXA) UNIV OF TEXAS.
XX
PI Saunders GF, Mars WM;
XX
DR WPI; 1991-361533/49.
DR P-PSDB; AAR15222.
XX
PS Disclosure; Fig 2; 13pp; English.
XX
CC The precise position of the CDS is unsure due to the 13 unidentified
CC nucleotides at the 5' end of the sequence. The most likely position
CC for initiation of translation occurs at nucleotide 94 and the open
CC reading frame following this point is claimed. The first 18 codons
CC of the claimed ORF (excluding the ATG) encode a sequence of amino
CC acids that resembles other leader sequences and their cleavage sites.
CC The mrs sequence was isolated as clone C-A3 (ATCC 39868) from a cDNA
CC library constructed from total RNA from a patient suffering from
CC chronic phase, Ph'-positive CML. The mrs is localised to human
CC chromosome 8, specifically to bands 8q21.1-23.
XX
SQ Sequence 498 BP; 118 A; 134 C; 122 G; 111 T; 13 other;

Alignment Scores:
Pred. No.:
Score: 0.0614
Percent Similarity: 80.00   Length: 498
Best Local Similarity: 55.17%   Matches: 13
Query Match: 44.83%   Conservative: 3
DB: 41.88%   Mismatches: 13
                    12   Indels: 0
                                   Gaps: 0

US-10-045-180A-6 (1-31) x AAQ14908 (1-498)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
         ||||||:|||||:||||:|||
Db 289 TGCTATTGCGAGAATACCAGCGTGCATTGCAGGAGAACGTGCCTATGGAACCTGCGATCTAC 348

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
         |||| ::
Db 349 CAGGGAAGACTCTGGGCCATTCTGCTGC 375

XX
PF 13-FEB-2002; 2002WO-JP01193.
XX
PR 14-FEB-2001; 2001JP-0036362.
XX
PA (GENO-) GENOX RES INC.
PA (NIGH-) JAPAN GEN AGENCY NATION.
XX
PI Sugita Y, Heishi M, Kagaya S, Gunji S, Saito H;
XX
DR WPI; 2002-674884/72.
DR P-PSDB; ABB98494.
XX
PS Disclosure; Page 45-46; 57pp; Japanese.
XX
CC The present invention relates to a method for examining allergic
CC diseases. The method comprises: (a) determining the expression level of
CC defensin 1 gene in the biological sample from a patient; and (b)
CC comparing the expression level with that in the sample of a healthy
CC individual. The method is for examining allergic diseases particularly
CC atopic dermatitis and its diagnosis, which is also applicable in
CC screening candidate compounds for remedies. The present sequence is the
CC coding sequence for human defensin 1.
XX
SQ Sequence 498 BP; 118 A; 134 C; 122 G; 111 T; 13 other;

Alignment Scores:
Pred. No.:
Score: 0.0614
Percent Similarity: 80.00   Length: 498
Best Local Similarity: 55.17%   Matches: 13
Query Match: 44.83%   Conservative: 3
DB: 41.88%   Mismatches: 13
                    24   Indels: 0
                                   Gaps: 0

US-10-045-180A-6 (1-31) x ABQ80601 (1-498)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
         ||||||:|||||:||||:|||
Db 289 TGCTATTGCGAGAATACCAGCGTGCATTGCAGGAGAACGTGCCTATGGAACCTGCGATCTAC 348

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
         |||| ::
Db 349 CAGGGAAGACTCTGGGCCATTCTGCTGC 375

XX
PN WO20026483Z-A1.
XX
FT CDS
FT   /tag= a
FT   /product= "Human defensin 1"
FH Key   Location/Qualifiers
FH CDS   94..378
XX
OS Homo sapiens.
XX
KW human; gene; ds.
KW Antiallergic; allergic disease; allergy; defensin 1; atopic dermatitis;
DE Human defensin 1 coding sequence.
XX
DT 11-NOV-2002 (first entry)
XX
AC ABQ80601;
ID ABQ80601 standard; DNA; 498 BP.
XX
RESULT 13
ABQ80601

XX
PN 9US-0122487.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PF 21-FEB-2000; 2000EP-0200610.
PD 06-SEP-2000.
XX
PD 2000EP-0200610.
XX
OS Homo sapiens.
XX
KW gene therapy; chromosome mapping; ss.
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
DE Human secreted protein 5' EST, SEQ ID NO: 96.
XX
DT 06-OCT-2000 (first entry)
XX
AC AAC00098;
ID AAC00098 standard; cDNA; 509 BP.
XX
RESULT 14
AAC00098

RESULT 15
ABN59722
ID   ABN59722 standard; cDNA, 514 BP.
XX
AC   ABN59722;
XX
DT   28-JUN-2002 (first entry)
XX
DE   Novel human coding sequence SEQ ID NO. 133.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2002226660-A2.
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US26015.
XX
PR   11-SEP-2000; 2000US-0659671.
XX
PA   (HYSE-) HYSEQ INC.

XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00092.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 96; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 509 BP; 118 A; 145 C; 132 G; 109 T; 5 other;

Alignment Scores:
Pred. No.:
Score: 0.063
Percent Similarity: 80.00   Length: 509
Best Local Similarity: 55.17%   Matches: 13
Query Match: 44.83%   Conservative: 3
DB: 41.88%   Mismatches: 13
                        21   Indels: 0
                                        Gaps: 0

US-10-045-180A-6 (1-31) x AAC00098 (1-509)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
         |||:::|||:::||||||||
Db 377 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTTCGCTATGGAACCTGCATCTAC 436

Qy 22 LeuGlyGluArgTyrProIleCysCys 30
          |||
Db 437 CAGGGAAGACTCTGGGCATTCTGCTGC 463

XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   P-PSDB; ABB97309.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Claim 1; SEQ ID NO 133; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic cell anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC   e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a coding sequence of the
CC   invention.
XX
SQ   Sequence 514 BP; 138 A; 142 C; 118 G; 116 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 0.0638
Percent Similarity: 80.00   Length: 514
Best Local Similarity: 55.17%   Matches: 13
Query Match: 44.83%   Conservative: 3
DB: 41.88%   Mismatches: 13
                        24   Indels: 0
                                        Gaps: 0

US-10-045-180A-6 (1-31) x ABN59722 (1-514)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
         |||:::|||:::||||||||
Db 292 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTTCGCTATGGAACCTGCATCTAC 351

Qy 22 LeuGlyGluArgTyrProIleCysCys 30
          |||
Db 352 CAGGGAAGACTCTGGGCATTCTGCTGC 378

Search completed: December 17, 2003, 15:06:05
Job time : 118.25 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: December 17, 2003, 15:06:18 ; Search time 107.181 Seconds
(without alignments)
963.776 Million cell updates/sec

Sequence: 1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31
Perfect score: 191
Title: US-10-045-180A-6

Scoring table: BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop 6.0 ; Delext 7.0

Searched: 2211978 seqs, 166101734 residues
Total number of hits satisfying chosen parameters: 4423956
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool_p/US10045180//runat_17122003_145026_12635/app_query.fasta_1.860
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10045180 @CGN 1_1.174 @runat_17122003_145026_12635
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US07_PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Result No.	Score	Query %	Length	Match	DB ID	Description
1	191	100.0	453	14	US-10-013-770-2	Sequence 2, Appli
2	191	100.0	453	14	US-10-045-180A-2	Sequence 2, Appli
3	191	100.0	4415	14	US-10-013-770-1	Sequence 1, Appli
4	191	100.0	4415	14	US-10-045-180A-1	Sequence 1, Appli
5	80	41.9	337	11	US-09-918-995-18198	Sequence 18198, A
6	80	41.9	337	11	US-09-918-995-19070	Sequence 19070, A
7	80	41.9	353	11	US-09-918-995-18323	Sequence 18323, A
8	80	41.9	555	13	US-10-252-157-223	Sequence 223, Appl
9	80	41.9	726	10	US-09-816-828-16	Sequence 16, Appl
10	80	41.9	726	10	US-09-816-828-16	Sequence 16, Appl
11	80	41.9	1348	15	US-10-125-237-78	Sequence 78, Appl
12	80	41.9	1348	15	US-10-105-891-78	Sequence 78, Appl
13	74	38.7	412	10	US-09-735-457-1047	Sequence 1047, Ap
14	74	38.7	412	10	US-09-902-941-1047	Sequence 1047, Ap
15	74	38.7	412	10	US-09-849-626-1047	Sequence 1047, Ap
16	74	38.7	412	13	US-10-111-872-1047	Sequence 1047, Ap
17	74	38.7	412	15	US-10-017-754-1047	Sequence 1047, Ap
18	74	38.7	967	13	US-10-162-043-6	Sequence 6, Appli
19	69	36.1	343	11	US-09-918-995-18788	Sequence 18788, A
20	67	35.1	588	13	US-10-029-386-18673	Sequence 18673, A
21	67	35.1	588	13	US-10-029-386-4917	Sequence 4917, Ap
22	66.5	34.8	640	13	US-10-027-632-106465	Sequence 106465,
23	66.5	34.8	640	13	US-10-027-632-106466	Sequence 106466,
24	66.5	34.8	640	13	US-10-027-632-106467	Sequence 106467,
25	66.5	34.8	640	14	US-10-027-632-106465	Sequence 106465,
26	66.5	34.8	640	14	US-10-027-632-106466	Sequence 106466,
27	66.5	34.8	640	14	US-10-027-632-106467	Sequence 106467,
28	66.5	34.8	738	13	US-10-027-632-22769	Sequence 22769, A
29	66.5	34.8	738	13	US-10-027-632-22770	Sequence 22770, A
30	66.5	34.8	738	13	US-10-027-632-22771	Sequence 22771, A
31	66.5	34.8	738	14	US-10-027-632-22769	Sequence 22769, A
32	66.5	34.8	738	14	US-10-027-632-22770	Sequence 22770, A
33	66.5	34.8	738	14	US-10-027-632-22771	Sequence 22771, A
34	66.5	34.8	542	14	US-10-045-180A-8	Sequence 8, Appli
35	65	34.0	784	13	US-10-027-632-165609	Sequence 165609,
36	65	34.0	784	14	US-10-027-632-165609	Sequence 165609,
37	65	34.0	787	13	US-10-027-632-25141	Sequence 25141, A
38	65	34.0	787	13	US-10-027-632-25142	Sequence 25142, A
39	65	34.0	787	13	US-10-027-632-25143	Sequence 25143, A
40	65	34.0	787	14	US-10-027-632-25141	Sequence 25141, A
41	65	34.0	787	14	US-10-027-632-25142	Sequence 25142, A
42	65	34.0	787	14	US-10-027-632-25143	Sequence 25143, A
43	65	34.0	860	14	US-10-044-090-473	Sequence 473, App
44	65	34.0	4295	14	US-10-045-180A-7	Sequence 7, Appli
45	63	33.0	15456	11	US-09-900-112-35	Sequence 35, Appl

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
US-10-013-770-2
Sequence 2, Application US/10013770
Publication No. US20020115151A1
GENERAL INFORMATION:
APPLICANT: GENSET SA
TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
THERAPEUTIC APPLICATIONS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 550 West C Street
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/013,770

FILING DATE: 10-Dec-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/486,580
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hart, Daniel
REGISTRATION NUMBER: 40,637
REFERENCE/DOCKET NUMBER: GENSET.064C1
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 453 BASE PAIRS
TYPE: NUCLEOTIDE
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-013-770-2

Alignment Scores:
Pred. No.:
Score: 4.36e-19   Length: 453
Percent Similarity: 191.00   Matches: 31
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 14   Indels: 0
                                   Gaps: 0

US-10-045-180A-6 (1-31) x US-10-013-770-2 (1-453)

```
Qy    1 IleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPhe  20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ARCTGCCATTGCAGAGTACTATACCTGCATTTTTGGAGAACATCTTGGTGGGACCTGCTTC 300

Qy   21 IleLeuGlyGluArgTyrProIleCysCysTyr 31
        |||||||||||||||||||||||||||||||||
Db  301 ATCCTTGGTGAACGCTACCCAATCTGCTGCTAC 333
```

RESULT 2
US-10-045-180A-2
Sequence 2, Application US/10045180A
Publication No. US20020182703A1
GENERAL INFORMATION:
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Composit
TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeuti
FILE REFERENCE: GEN-1001I
CURRENT APPLICATION NUMBER: US/10/045,180A
CURRENT FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: US 09/486,580
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: PCT/FR98/01864
PRIOR FILING DATE: 1998-08-28
PRIOR APPLICATION NUMBER: FR 97/10823
PRIOR FILING DATE: 1997-08-29
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patentin version 3.1
SEQ ID NO 2
LENGTH: 453
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (52)..(336)
OTHER INFORMATION: Def-X coding sequence
US-10-045-180A-2

Alignment Scores:
Pred. No.:
Score: 4.36e-19   Length: 453
Percent Similarity: 191.00   Matches: 31
Best Local Similarity: 100.00%   Conservative: 0
                                   Mismatches: 0

Query Match: 100.00%   Indels: 0
DB: 14   Gaps: 0

US-10-045-180A-6 (1-31) x US-10-045-180A-2 (1-453)

```
Qy    1 IleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPhe  20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ARCTGCCATTGCAGAGTACTATACCTGCATTTTTGGAGAACATCTTGGTGGGACCTGCTTC 300

Qy   21 IleLeuGlyGluArgTyrProIleCysCysTyr 31
        |||||||||||||||||||||||||||||||||
Db  301 ATCCTTGGTGAACGCTACCCAATCTGCTGCTAC 333
```

RESULT 3
US-10-013-770-1
Sequence 1, Application US/10013770
Publication No. US20020115151A1
GENERAL INFORMATION:
APPLICANT: GENSET SA
TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
   COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
   THERAPEUTIC APPLICATIONS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 550 West C Street
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/013,770
FILING DATE: 10-Dec-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/486,580
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hart, Daniel
REGISTRATION NUMBER: 40,637
REFERENCE/DOCKET NUMBER: GENSET.064C1
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4415 BASE PAIRS
TYPE: NUCLEOTIDE
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Exon 1
LOCATION: 1836..1874
FEATURE:
NAME/KEY: Exon 2
LOCATION: 3394..3577
FEATURE:
NAME/KEY: Exon 3
LOCATION: 4161..4380
FEATURE:
NAME/KEY: start CDS
LOCATION: 3406..3408
FEATURE:
NAME/KEY: stop CDS
LOCATION: 4276..4278
FEATURE:
NAME/KEY: polyadenylation site
LOCATION: 4374..4379
SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-10-013-770-1

Alignment Scores:

Pred. No.:

Score:	7.29e-18	Length:	4415
Percent Similarity:	191.00	Matches:	31
Best Local Similarity:	100.00%	Conservative:	0
Query Match:	100.00%	Mismatches:	0
DB:	100.00%	Indels:	0
	14	Gaps:	0

US-10-045-180A-6 (1-31) x US-10-013-770-1 (1-4415)

```
Qy     1 IleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPhe 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4181 ATCTGCCATTGCAGAGTACTATATACTGCATTTTTGGAGAACATCTTGGTGGGACCTGCTTC 4240

Qy    21 IleLeuGlyGluArgTyrTyrProIleCysCysTyr 31
           ||||||||||||||||||||||||||||||||||||
Db  4241 ATCCTTGGTGAACGGCTACCCAATCTGCTGCTAC 4273
```

RESULT 4
US-10-045-180A-1

; Sequence 1, Application US/10045180A
; Publication No. US20020182703A1
; GENERAL INFORMATION:
; APPLICANT: Bouguelaret, Lydie
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Compositi
; TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeutic
; FILE REFERENCE: GEN-100D1
; CURRENT APPLICATION NUMBER: US/10/045,180A
; CURRENT FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 09/486,580
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: PCT/FR98/01864
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: FR 97/10823
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
;
; SEQ ID NO 1
; LENGTH: 4415
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(4415)
; OTHER INFORMATION: Def-X genomic sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (85)..(85)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (143)..(143)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (670)..(670)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (970)..(970)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1111)..(1111)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1150)..(1150)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:

; FEATURE:
; NAME/KEY: CAAT signal
; LOCATION: (1711)..(1714)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: TATA_signal
; LOCATION: (1758)..(1767)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1717)..(1717)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1875)..(1880)
; OTHER INFORMATION: splice donor site
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1780)..(1780)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1974)..(1974)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1836)..(1874)
; OTHER INFORMATION: Exon 1
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2117)..(2117)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2133)..(2133)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2155)..(2335)
; OTHER INFORMATION: Alu insertion
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2186)..(2186)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2191)..(2191)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2367)..(2367)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2710)..(2780)
; OTHER INFORMATION: L1 fragment insertion
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3391)..(3393)
; OTHER INFORMATION: splice acceptor site
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3394)..(3577)
; OTHER INFORMATION: Exon 2
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3406)..(3408)
; OTHER INFORMATION: Translation initiation codon (ATG)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3578)..(3583)
; OTHER INFORMATION: splice donor site
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4123)..(4123)
; OTHER INFORMATION: n = a, c, g, or t.
; FEATURE:
; NAME/KEY: misc_feature
```

LOCATION: (4161)..(4163)
OTHER INFORMATION:
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4164)..(4379)
OTHER INFORMATION:
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4274)..(4276)
OTHER INFORMATION: Translation termination codon (TAA)
FEATURE:
NAME/KEY: polyA_signal
LOCATION: (4374)..(4379)
OTHER INFORMATION:
US-10-045-180A-1

Alignment Scores:
Pred. No.:
Score:                    191.00
Percent Similarity:       100.00%
Best Local Similarity:    100.00%
Query Match:              14

Qy    1 IleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyHisGlyThrCysPhe 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4181 ATCTGCCATTGCAGAGTACTATACTGCATTTTTGGAGAACACCTTGGTGTGGGACCTGCTTC 4240

Qy   21 IleLeuGlyGluArgTyrProIleCysCysTyr 31
          |||||||||||||||||||||||||||||||||
Db 4241 ATCCTTGGTGAACGCTACCCAATCTGCTGCTAC 4273

RESULT 5
US-09-918-995-18198/c
Sequence 18198, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 18198
LENGTH: 337
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(337)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-18198

Alignment Scores:
Pred. No.:
Score:                    0.00788
Percent Similarity:       80.00
Best Local Similarity:    55.17%
Query Match:              44.83%
                          41.88%

US-10-045-180A-6 (1-31) x US-09-918-995-18198 (1-337)

Qy    2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyHisGlyThrCysPheIle 21
          ||:|:||||||||:|||||| |||||| |||||:|||
Db  198 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTGCGTTATGGAACCTGCCATCTAC 139

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
          |||:|
Db  138 CAGGGAAGAGACTCTGGGCAATTCTGCTGC 112

RESULT 6
US-09-918-995-19070/c
Sequence 19070, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
SEQ ID NO 19070
LENGTH: 337
TYPE: DNA
ORGANISM: Homo sapiens
US-09-918-995-19070

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
          ||:||
Db  138 CAGGGAAGAGACTCTGGGCAATTCTGCTGC 112

Alignment Scores:
Pred. No.:
Score:                    0.00788
Percent Similarity:       80.00
Best Local Similarity:    55.17%
Query Match:              44.83%
                          41.88%

US-10-045-180A-6 (1-31) x US-09-918-995-19070 (1-337)

Qy    2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyHisGlyThrCysPheIle 21
          ||:|:||||||||:|||||| |||||| |||||:|||
Db  198 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTGCGTTATGGAACCTGCCATCTAC 139

RESULT 7
US-09-918-995-18323
Sequence 18323, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 18323
LENGTH: 353
TYPE: DNA
ORGANISM: Homo sapiens
US-09-918-995-18323

Alignment Scores:
Pred. No.:
Score:                    0.00835
Percent Similarity:       80.00
Best Local Similarity:    55.17%
Query Match:              44.83%
                          41.88%

US-10-045-180A-6 (1-31) x US-09-918-995-18323 (1-353)

```
Qy   2  CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle  21
        |:::||||||:|||||
Db 147  TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTATGGAACCTGCATCTAC 206

Qy  22  LeuGlyGluArgTyrProIleCysCys  30
        ||||
Db 207  CAGGGAAGACTCTGGGCATTCTGCTGC 233
```

RESULT 8
US-10-252-157-223
Sequence 223, Application US/10252157
Publication No. US20030190640A1
GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Pearson, Cecelia I.
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252,157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 223
; LENGTH: 555
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030190640A1 903338.12
US-10-252-157-223

Alignment Scores:
```
Pred. No.:                              Length:        555
Score:                    0.0146        Matches:       13
Percent Similarity:       80.00         Conservative:  3
Best Local Similarity:    55.17%        Mismatches:    13
Query Match:              44.83%        Indels:        0
DB:                       41.88%        Gaps:          0
                          13
```

US-10-045-180A-6 (1-31) x US-10-252-157-223 (1-555)

```
Qy   2  CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle  21
        |:::||||||:|||||
Db 340  TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTATGGAACCTGCATCTAC 399

Qy  22  LeuGlyGluArgTyrProIleCysCys  30
        ||||
Db 400  CAGGGAAGACTCTGGGCATTCTGCTGC 426
```

RESULT 9
US-09-816-828-16
Sequence 16, Application US/09816828
Patent No. US20020150898A1
GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhao, Qing A.
; APPLICANT: Zhang, Jie
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020150898A1e1 Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 791CIP2E
; CURRENT APPLICATION NUMBER: US/09/816,828
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/770,160
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/552,929
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 16
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (151)..(522)
US-09-816-828-16

Alignment Scores:
```
Pred. No.:                              Length:        726
Score:                    0.0204        Matches:       13
Percent Similarity:       80.00         Conservative:  3
Best Local Similarity:    55.17%        Mismatches:    13
Query Match:              44.83%        Indels:        0
DB:                       41.88%        Gaps:          0
                          10
```

US-10-045-180A-6 (1-31) x US-09-816-828-16 (1-726)

```
Qy   2  CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle  21
        |:::||||||:|||||
Db 436  TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTATGGAACCTGCATCTAC 495

Qy  22  LeuGlyGluArgTyrProIleCysCys  30
        ||||
Db 496  CAGGGAAGACTCTGGGCATTCTGCTGC 522
```

RESULT 10
US-09-816-828-16/c
Sequence 16, Application US/09816828
Patent No. US20020150898A1
GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhao, Qing A.
; APPLICANT: Zhang, Jie
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020150898A1e1 Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 791CIP2E
; CURRENT APPLICATION NUMBER: US/09/816,828
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/770,160
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/552,929
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 16
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (151)..(522)
US-09-816-828-16

Alignment Scores:
```
Pred. No.:                              Length:    726
Score:                    0.0204        Matches:   13
```

Percent Similarity: 55.17%   Conservative: 3
Best Local Similarity: 44.83%   Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: Gaps: 0

US-10-045-180A-6 (1-31) x US-09-816-828-16 (1-726)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
          |||::|||||||||::| |||||||
Db   100 TGCTATTGCAGAATACCAGCGGTGCATTGCAGGAGAACGTCGCTATGGAACCTGCATCTAC 41

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
          ||| ::: |||||
Db   40 CAGGGAAGACTCTGGGCATTCTGCTGC 14

RESULT 11
US-10-125-237-78
; Sequence 78, Application US/10125237
; Publication No. US20030022329A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhao, Qing A.
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhang, Jie
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030022329A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 791CIP2ADIV
; CURRENT APPLICATION NUMBER: US/10/125,237
; CURRENT FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 09/668,317
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/552,929
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 78
; LENGTH: 1348
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(1143)
US-10-125-237-78

Alignment Scores:
Pred. No.:   0.0438   Length: 1348
Score: 80.00   Matches: 13
Percent Similarity: 55.17%   Conservative: 3
Best Local Similarity: 44.83%   Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: 15   Gaps: 0

US-10-045-180A-6 (1-31) x US-10-125-237-78 (1-1348)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
          |||::|||||||||::| |||||||
Db   1054 TGCTATTGCAGAATACCAGCGGTGCATTGCAGGAGAACGTCGCTATGGAACCTGCATCTAC 1113

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
          ||| ::: |||||
Db   1114 CAGGGAAGACTCTGGGCATTCTGCTGC 1140

RESULT 12
US-10-105-891-78
; Sequence 78, Application US/10105891
; Publication No. US20030073099A1

GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhao, Qing A.
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhang, Jie
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030073099A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 791CIP2A
; CURRENT APPLICATION NUMBER: US/10/105,891
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: 09/668,317
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/552,929
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 78
; LENGTH: 1348
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1143)
US-10-105-891-78

Alignment Scores:
Pred. No.:   0.0438   Length: 1348
Score: 80.00   Matches: 13
Percent Similarity: 55.17%   Conservative: 3
Best Local Similarity: 44.83%   Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: Gaps: 0

US-10-045-180A-6 (1-31) x US-10-105-891-78 (1-1348)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
          |||::|||||||||::| |||||||
Db   1054 TGCTATTGCAGAATACCAGCGGTGCATTGCAGGAGAACGTCGCTATGGAACCTGCATCTAC 1113

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
          ||| ::: |||||
Db   1114 CAGGGAAGACTCTGGGCATTCTGCTGC 1140

RESULT 13
US-09-736-457-1047/c
; Sequence 1047, Application US/09736457
; Patent No. US2002016637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1047

```
;   LENGTH: 412
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(412)
;   OTHER INFORMATION: n = A,T,C or G
US-09-736-457-1047

Alignment Scores:
Pred. No.:                      0.078       Length:          412
Score:                          74.00       Matches:          13
Percent Similarity:             51.72%      Conservative:      2
Best Local Similarity:          44.83%      Mismatches:       14
Query Match:                    38.74%      Indels:            0
DB:                             10          Gaps:              0

US-10-045-180A-6 (1-31) x US-09-736-457-1047 (1-412)

Qy       2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
           ||::|||||         ||  ||   ||||||||
Db     254 TGCTATTGCCGAACCGGCCGTTGTGCTACCCGTGAGTCCCTCTCCGGGGTGTGTGAAATC 195

Qy      22 LeuGlyGluArgTyrProIleCysCys 30
           ||||||||           :::|||||||
Db     194 AGTGGCCGCCCTNTACAGACTCTGCTGT 168
```

RESULT 14

```
US-09-902-941-1047/c
;   Sequence 1047, Application US/09902941
;   Patent No. US20020172952A1
;   GENERAL INFORMATION:
;   APPLICANT: Henderson, Robert A.
;   APPLICANT: Wang, Tongtong
;   APPLICANT: Watanabe, Yoshihiro
;   APPLICANT: Johnson, Jeffrey C.
;   APPLICANT: Retter, Marc W.
;   APPLICANT: Marnerakis, Margarita
;   APPLICANT: Carter, Darrick
;   APPLICANT: Fanger, Gary R.
;   APPLICANT: Vedvick, Thomas S.
;   APPLICANT: Bangur, Chaitanya S.
;   APPLICANT: McNabb, Andria
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;   TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;   FILE REFERENCE: 210121.478C17
;   CURRENT APPLICATION NUMBER: US/09/902,941
;   CURRENT FILING DATE: 2001-07-10
;   NUMBER OF SEQ ID NOS: 2002
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 1047
;   LENGTH: 412
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 183, 271, 287, 292, 294, 343
;   OTHER INFORMATION: n = A,T,C or G
US-09-902-941-1047

Alignment Scores:
Pred. No.:                      0.078       Length:          412
Score:                          74.00       Matches:          13
Percent Similarity:             51.72%      Conservative:      2
Best Local Similarity:          44.83%      Mismatches:       14
Query Match:                    38.74%      Indels:            0
DB:                             10          Gaps:              0

US-10-045-180A-6 (1-31) x US-09-902-941-1047 (1-412)

Qy       2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
           ||::|||||         ||  ||   ||||||||
Db     254 TGCTATTGCCGAACCGGCCGTTGTGCTACCCGTGAGTCCCTCTCCGGGGTGTGTGAAATC 195
```

```
Qy      22 LeuGlyGluArgTyrProIleCysCys 30
           ||||||||           :::|||||||
Db     194 AGTGGCCGCCCTNTACAGACTCTGCTGT 168
```

RESULT 15

```
US-09-849-626-1047/c
;   Sequence 1047, Application US/09849626
;   Publication No. US2002019766A1
;   GENERAL INFORMATION:
;   APPLICANT: Bangur, Chaitanya
;   APPLICANT: Fanger, Gary
;   APPLICANT: Wang, Aijun
;   APPLICANT: Wang, Tongtong
;   APPLICANT: Switzer, Anne
;   APPLICANT: McNeill, Patricia
;   APPLICANT: Clapper, Jonathan
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;   TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
;   FILE REFERENCE: 210121.478C16
;   CURRENT APPLICATION NUMBER: US/09/849,626
;   CURRENT FILING DATE: 2001-05-03
;   NUMBER OF SEQ ID NOS: 1926
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 1047
;   LENGTH: 412
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(412)
;   OTHER INFORMATION: n = A,T,C or G
US-09-849-626-1047

Alignment Scores:
Pred. No.:                      0.078       Length:          412
Score:                          74.00       Matches:          13
Percent Similarity:             51.72%      Conservative:      2
Best Local Similarity:          44.83%      Mismatches:       14
Query Match:                    38.74%      Indels:            0
DB:                             10          Gaps:              0

US-10-045-180A-6 (1-31) x US-09-849-626-1047 (1-412)

Qy       2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
           ||::|||||         ||  ||   ||||||||
Db     254 TGCTATTGCCGAACCGGCCGTTGTGCTACCCGTGAGTCCCTCTCCGGGGTGTGTGAAATC 195

Qy      22 LeuGlyGluArgTyrProIleCysCys 30
           ||||||||           :::|||||||
Db     194 AGTGGCCGCCCTNTACAGACTCTGCTGT 168
```

```
Search completed: December 17, 2003, 21:36:03
Job time : 109.181 secs
```

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:     December 17, 2003, 14:50:33 ; Search time 31 Seconds
                                (without alignments)
                                441.383 Million cell updates/sec

Title:      US-10-045-180A-6
Perfect score:   191
Sequence:   1 ICHCRVLYCIFGEHLGGTCPILGBRYPICCY 31

Scoring table:  BLOSUM62
            Xgapop  10.0 , Xgapext   0.5
            Ygapop  10.0 , Ygapext   0.5
            Fgapop   6.0 , Fgapext   7.0
            Delop    6.0 , Delext    7.0

Searched:    569978 seqs, 22069l566 residues

Total number of hits satisfying chosen parameters:    1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/UsPTO_spool_p/US10045180/runat 17122003 145024 12553/app query.fasta_1.860
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10045180 @CGN_1_1 193 @runat 17122003 145024 12553 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Issued_Patents_NA:*
             1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
             2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
             3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
             4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
             5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
             6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	191	100.0	453	4	US-09-486-580A-2	Sequence 2, Appli
2	191	100.0	4415	4	US-09-486-580A-1	Sequence 1, Appli
3	80	41.9	90	3	US-09-230-180-27	Sequence 27, Appl
4	80	41.9	108	1	US-08-282-030-43	Sequence 43, Appl
c 5	80	41.9	108	1	US-08-282-030-44	Sequence 44, Appl
6	80	41.9	108	1	US-08-282-030-45	Sequence 45, Appl
7	80	41.9	108	5	PCT-US95-10219-43	Sequence 43, Appl
c 8	80	41.9	108	5	PCT-US95-10219-44	Sequence 44, Appl
9	80	41.9	108	5	PCT-US95-10219-45	Sequence 45, Appl
c 10	80	41.9	109	1	US-08-282-030-61	Sequence 61, Appl
c 11	80	41.9	109	5	PCT-US95-10219-61	Sequence 61, Appl
c 12	80	41.9	1097	1	US-08-021-608D-5	Sequence 5, Appli
c 13	80	41.9	1097	1	US-08-726-160-5	Sequence 5, Appli
c 14	80	41.9	1097	5	PCT-US94-01782-5	Sequence 5, Appli
15	74	38.7	123	1	US-08-282-030-50	Sequence 50, Appl
c 16	74	38.7	123	1	US-08-282-030-62	Sequence 62, Appl
17	74	38.7	123	5	PCT-US95-10219-50	Sequence 50, Appl
c 18	74	38.7	123	5	PCT-US95-10219-62	Sequence 62, Appl
c 19	74	38.7	412	4	US-09-702-705-1047	Sequence 1047, Ap
c 20	74	38.7	412	4	US-09-736-457-1047	Sequence 1047, Ap
21	74	38.7	424	1	US-08-158-189-4	Sequence 4, Appli
22	74	38.7	967	4	US-09-450-972-6	Sequence 6, Appli
23	74	38.7	2880	1	US-08-158-189-1	Sequence 1, Appli
24	69.5	36.4	365	1	US-08-486-013-43	Sequence 43, Appl
25	69.5	36.4	365	2	US-08-482-279-43	Sequence 43, Appl
26	69.5	36.4	365	2	US-08-342-268-43	Sequence 43, Appl
27	69.5	36.4	365	3	US-09-015-968-43	Sequence 43, Appl
28	69.5	36.4	365	4	US-09-397-386-43	Sequence 43, Appl
29	69	36.1	420	1	US-08-486-013-46	Sequence 46, Appl
30	69	36.1	420	2	US-08-482-279-46	Sequence 46, Appl
31	69	36.1	420	2	US-08-342-268-46	Sequence 46, Appl
32	69	36.1	420	3	US-09-015-968-46	Sequence 46, Appl
33	69	36.1	420	4	US-09-397-386-46	Sequence 46, Appl
34	67	35.1	342	1	US-08-486-013-47	Sequence 47, Appl
35	67	35.1	342	2	US-08-482-279-47	Sequence 47, Appl
36	67	35.1	342	2	US-08-342-268-47	Sequence 47, Appl
37	67	35.1	342	3	US-09-015-968-47	Sequence 47, Appl
38	67	35.1	342	3	US-09-397-386-47	Sequence 47, Appl
39	67	35.1	388	1	US-08-486-013-52	Sequence 52, Appl
40	67	35.1	388	2	US-08-482-279-52	Sequence 52, Appl
41	67	35.1	388	2	US-08-342-268-52	Sequence 52, Appl
42	67	35.1	388	3	US-09-015-968-52	Sequence 52, Appl
43	67	35.1	388	4	US-09-397-386-52	Sequence 52, Appl
44	66.5	34.8	452	1	US-08-158-189-6	Sequence 6, Appli
45	66.5	34.8	452	4	US-09-016-434-1198	Sequence 1198, Ap

ALIGNMENTS

```
RESULT 1
US-09-486-580A-2
; Sequence 2, Application US/09486580A
; Patent No. 6329340
; GENERAL INFORMATION:
;  APPLICANT: GENSET SA
;  TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;  TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;  NUMBER OF SEQUENCES: 6
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/486,580A
;   FILING DATE: FEBRUARY 25, 2000
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
;  INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH: 453 BASE PAIRS
;    TYPE: NUCLEOTIDE
;    STRANDEDNESS: DOUBLE
;    TOPOLOGY: LINEAR
;    MOLECULE TYPE: cDNA
```

; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-486-580A-2

Alignment Scores:
Pred. No.: 3.5e-19
Score: 191.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 4

Length: 453
Matches: 31
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-045-180A-6 (1-31) x US-09-486-580A-2 (1-453)

```
Qy   1 IleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPhe 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 ATCTGCCATTGCAGAGTACTATACTGCATTTTTGGAGAACATCTTGGTGGGACCTGCTTC 300

Qy  21 IleLeuGlyGluArgTyrProIleCysCysTyr 31
          |||||||||||||||||||||||||||||||||
Db 301 ATCCTTGGTGAACGCTACCCAATCTGCTGCTAC 333
```

RESULT 2
US-09-486-580A-1
; Sequence 1, Application US/09486580A
; Patent No. 6329340
; GENERAL INFORMATION:
;   APPLICANT: GENSET SA
;   TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;   TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;   TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/486,580A
;   FILING DATE: FEBRUARY 25, 2000
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 4415 BASE PAIRS
;   TYPE: NUCLEOTIDE
;   STRANDEDNESS: DOUBLE
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE: DNA
;   ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: Exon 1
;   LOCATION: 1836..1874
;   FEATURE:
;   NAME/KEY: Exon 2
;   LOCATION: 3394..3577
;   FEATURE:
;   NAME/KEY: Exon 3
;   LOCATION: 4161..4380
;   FEATURE:
;   NAME/KEY: start CDS
;   LOCATION: 3406..3408
;   FEATURE:
;   NAME/KEY: stop CDS
;   LOCATION: 4276..4278
;   FEATURE:
;   NAME/KEY: polyadenylation site
;   LOCATION: 4374..4379
US-09-486-580A-1

Alignment Scores:
Pred. No.: 6.14e-18
Score: 191.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 4

Length: 4415
Matches: 31
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-045-180A-6 (1-31) x US-09-486-580A-1 (1-4415)

```
Qy    1 IleCysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPhe 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4181 ATCTGCCATTGCAGAGTACTATACTGCATTTTTGGAGAACATCTTGGTGGGACCTGCTTC 4240

Qy   21 IleLeuGlyGluArgTyrProIleCysCysTyr 31
           |||||||||||||||||||||||||||||||||
Db 4241 ATCCTTGGTGAACGCTACCCAATCTGCTGCTAC 4273
```

RESULT 3
US-09-230-180-27
; Sequence 27, Application US/09230180
; Patent No. 6183992
; GENERAL INFORMATION:
;   APPLICANT: Kim, Sun-Chang
;   APPLICANT: Lee, Jae Hyun
;   APPLICANT: Kang, Min Hyung
;   APPLICANT: Kim, Jeong Hyun
;   APPLICANT: Hong, Seung-Suh
;   APPLICANT: Lee, Hyun-Soo
;   APPLICANT: Samyang Genex Corporation
;   APPLICANT: Korea Advanced Institute of Science and Technology
;   TITLE OF INVENTION: METHOD FOR MASS PRODUCTION OF
;   TITLE OF INVENTION: ANTIMICROBIAL PEPTIDE
;   FILE REFERENCE: 6181/0F135
;   CURRENT APPLICATION NUMBER: US/09/230,180
;   CURRENT FILING DATE: 1999-03-10
;   PRIOR APPLICATION NUMBER: PCT/KR98/00132
;   PRIOR FILING DATE: 1998-05-28
;   PRIOR APPLICATION NUMBER: KR 13372/1998
;   PRIOR FILING DATE: 1998-04-09
;   PRIOR APPLICATION NUMBER: KR 21312/1997
;   PRIOR FILING DATE: 1997-05-28
;   NUMBER OF SEQ ID NOS: 36
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 27
;   LENGTH: 90
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: DNA sequence deduced from HNP-I peptide sequence
;   OTHER INFORMATION: based on codon usage of E. coli
US-09-230-180-27

Alignment Scores:
Pred. No.: 0.00058
Score: 80.00
Percent Similarity: 55.17%
Best Local Similarity: 44.83%
Query Match: 41.88%
DB: 3

Length: 90
Matches: 13
Conservative: 3
Mismatches: 13
Indels: 0
Gaps: 0

US-10-045-180A-6 (1-31) x US-09-230-180-27 (1-90)

```
Qy  2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
        ||:::|||||||:::     |||||   ||||       ||||||| |||||||||
Db  4 TGCTACTGCCGTATCCCGGCGTGCATCCGGGCGTGCATCGGCGGTGAGCGGTTACGGTACCTGCGTTGCGTCGATCTAC 63
```

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
            ||||||
Db   64 CAGGGTCGTCGTGTGGGGTTCTGCTGC 90

RESULT 4
US-08-282-030-43
; Sequence 43, Application US/08282030
; Patent No. 5589364
; GENERAL INFORMATION:
;   APPLICANT: Williams, Jon I.
;   APPLICANT: Pierce, James C.
;   APPLICANT: Anderson, Mark G.
;   APPLICANT: Kari, Prasad
;   TITLE OF INVENTION: Recombinant Production of Biologically
;   TITLE OF INVENTION: Active Peptides and Proteins
;   NUMBER OF SEQUENCES: 62
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;   ADDRESSEE: Dunner
;   STREET: 1300 I Street, N.W.
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20005-3315
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/282,030
;   FILING DATE: 29-JUL-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Fordis, Jean B.
;   REGISTRATION NUMBER: 32,984
;   REFERENCE/DOCKET NUMBER: 05387.0001-00000
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-408-4000
;   TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 43:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 108 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-282-030-43

Alignment Scores:
Pred. No.:   0.00073   Length: 108
Score: 80.00   Matches: 13
Percent Similarity: 55.17%   Conservative: 3
Best Local Similarity: 44.83%   Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: 1   Gaps: 0

US-10-045-180A-6 (1-31) x US-08-282-030-43 (1-108)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
           ||:::|||||||:::| ||||||
Db   12 TGTTACTGCCGTATTCCGGGCATGCATCGCACGGCGAGCGTCGCTATGGTACTTGTATTTAC 71

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
           ||| |||
Db   72 CAGGGTCGTCGTGTGGGGCATTCTGTTGC 98

RESULT 5
US-08-282-030-44/c
; Sequence 44, Application US/08282030
; Patent No. 5589364
; GENERAL INFORMATION:
;   APPLICANT: Williams, Jon I.
;   APPLICANT: Pierce, James C.
;   APPLICANT: Anderson, Mark G.
;   APPLICANT: Kari, Prasad
;   TITLE OF INVENTION: Recombinant Production of Biologically
;   TITLE OF INVENTION: Active Peptides and Proteins
;   NUMBER OF SEQUENCES: 62
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;   ADDRESSEE: Dunner
;   STREET: 1300 I Street, N.W.
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20005-3315
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/282,030
;   FILING DATE: 29-JUL-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Fordis, Jean B.
;   REGISTRATION NUMBER: 32,984
;   REFERENCE/DOCKET NUMBER: 05387.0001-00000
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-408-4000
;   TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 44:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 108 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-282-030-44

Alignment Scores:
Pred. No.:   0.00073   Length: 108
Score: 80.00   Matches: 13
Percent Similarity: 55.17%   Conservative: 3
Best Local Similarity: 44.83%   Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: 1   Gaps: 0

US-10-045-180A-6 (1-31) x US-08-282-030-44 (1-108)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
           ||:::|||||||:::| ||||||
Db 101 TGTTACTGCCGTATTCCGGGCATGCATCGCACGGCGAGCGTCGCTATGGTACTTGTATTTAC 42

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
           ||| |||
Db   41 CAGGGTCGTCGTGTGGGCATTCTGTTGC 15

RESULT 6
US-08-282-030-45
; Sequence 45, Application US/08282030
; Patent No. 5589364
; GENERAL INFORMATION:
;   APPLICANT: Williams, Jon I.
;   APPLICANT: Pierce, James C.
;   APPLICANT: Anderson, Mark G.
;   APPLICANT: Kari, Prasad
;   TITLE OF INVENTION: Recombinant Production of Biologically
;   TITLE OF INVENTION: Active Peptides and Proteins
;   NUMBER OF SEQUENCES: 62
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;   ADDRESSEE: Dunner
;   STREET: 1300 I Street, N.W.
;   CITY: Washington

```
;       STATE: D.C.
;       COUNTRY: USA
;       ZIP: 20005-3315
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/282,030
;       FILING DATE: 29-JUL-1994
;       CLASSIFICATION: 435
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Fordis, Jean B.
;       REGISTRATION NUMBER: 32,984
;       REFERENCE/DOCKET NUMBER: 05387.0001-00000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 202-408-4000
;       TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 45:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 108 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
US-08-282-030-45

Alignment Scores:
Pred. No.:               0.00073      Length:           108
Score:                   80.00        Matches:          13
Percent Similarity:      55.17%       Conservative:     3
Best Local Similarity:   44.83%       Mismatches:       13
Query Match:             41.88%       Indels:           0
DB:                      1            Gaps:             0

US-10-045-180A-6 (1-31) x US-08-282-030-45 (1-108)

Qy    2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
          ||:::||||||:::  |||||   ||||||
Db   12 TGTTACTGCCGTATTCCGGCATGCATCCGGCATGCGAGGCGAGCGGTCGCTATGGTACTTGTATTTAC 71

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
          |||||     :::
Db   72 CAGGGTCGTCGTGGGCATTCTGTTGC 98

RESULT 7
PCT-US95-10219-43
; Sequence 43, Application PC/TUS9510219
; GENERAL INFORMATION:
;     APPLICANT: Williams, Jon I.
;     APPLICANT: Pierce, James C.
;     APPLICANT: Anderson, Mark G.
;     APPLICANT: Karl, Prasad
;     TITLE OF INVENTION: Recombinant Production of Biologically
;     TITLE OF INVENTION: Active Peptides and Proteins
;     NUMBER OF SEQUENCES: 62
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;       ADDRESSEE: Dunner
;       STREET: 1300 I Street, N.W.
;       CITY: Washington
;       STATE: D.C.
;       COUNTRY: USA
;       ZIP: 20005-3315
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: PCT/US95/10219
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/282,030
;       FILING DATE: 29-JUL-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Fordis, Jean B.
;       REGISTRATION NUMBER: 32,984
;       REFERENCE/DOCKET NUMBER: 05387.0001-00000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 202-408-4000
;       TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 43:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 108 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
PCT-US95-10219-43

Alignment Scores:
Pred. No.:               0.00073      Length:           108
Score:                   80.00        Matches:          13
Percent Similarity:      55.17%       Conservative:     3
Best Local Similarity:   44.83%       Mismatches:       13
Query Match:             41.88%       Indels:           0
DB:                      5            Gaps:             0

US-10-045-180A-6 (1-31) x PCT-US95-10219-43 (1-108)

Qy    2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
          ||:::||||||:::  |||||   ||||||
Db   12 TGTTACTGCCGTATTCCGGCATGCATCCGGCATGCGAGGCGAGCGGTCGCTATGGTACTTGTATTTAC 71

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
          |||||     :::
Db   72 CAGGGTCGTCGTGGGCATTCTGTTGC 98

RESULT 8
PCT-US95-10219-44/c
; Sequence 44, Application PC/TUS9510219
; GENERAL INFORMATION:
;     APPLICANT: Williams, Jon I.
;     APPLICANT: Pierce, James C.
;     APPLICANT: Anderson, Mark G.
;     APPLICANT: Karl, Prasad
;     TITLE OF INVENTION: Recombinant Production of Biologically
;     TITLE OF INVENTION: Active Peptides and Proteins
;     NUMBER OF SEQUENCES: 62
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;       ADDRESSEE: Dunner
;       STREET: 1300 I Street, N.W.
;       CITY: Washington
;       STATE: D.C.
;       COUNTRY: USA
;       ZIP: 20005-3315
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: PCT/US95/10219
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/282,030
;       FILING DATE: 29-JUL-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Fordis, Jean B.
;       REGISTRATION NUMBER: 32,984
;       REFERENCE/DOCKET NUMBER: 05387.0001-00000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 202-408-4000
```

TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US95/10219-44

Alignment Scores:
Pred. No.: 0.00073   Length: 108
Score: 80.00   Matches: 13
Percent Similarity: 55.17% Conservative: 3
Best Local Similarity: 44.83% Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: 5   Gaps: 0

US-10-045-180A-6 (1-31) x PCT-US95-10219-44 (1-108)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
Db 101 TGTTACTGCCGTATTCCGGCATCATCGCAGGCGAGCGTCGCTATGGTACTTGTATTTAC 42

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
Db   41 CAGGGTCGTCTGTGGGCATTCTGTTGC 15

RESULT 9
PCT-US95-10219-45
; Sequence 45, Application PC/TUS9510219
; GENERAL INFORMATION:
;   APPLICANT: Williams, Jon I.
;   APPLICANT: Pierce, James C.
;   APPLICANT: Anderson, Mark G.
;   APPLICANT: Kari, Prasad
;   TITLE OF INVENTION: Recombinant Production of Biologically
;   TITLE OF INVENTION: Active Peptides and Proteins
;   NUMBER OF SEQUENCES: 62
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;   ADDRESSEE: Dunner
;   STREET: 1300 I Street, N.W.
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20005-3315
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/10219
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/282,030
;   FILING DATE: 29-JUL-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Fordis, Jean B.
;   REGISTRATION NUMBER: 32,984
;   REFERENCE/DOCKET NUMBER: 05387.0001-00000
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-408-4000
;   TELEFAX: 202-408-4400
;   INFORMATION FOR SEQ ID NO: 45:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 108 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
PCT-US95-10219-45

Alignment Scores:
Pred. No.: 0.00073   Length: 108
Score: 80.00   Matches: 13
Percent Similarity: 55.17% Conservative: 3
Best Local Similarity: 44.83% Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: 5   Gaps: 0

US-10-045-180A-6 (1-31) x PCT-US95-10219-45 (1-108)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
Db   12 TGTTACTGCCGTATTCCGGCATGCATCGCAGGCGAGCGTCGCTATGGTACTTGTATTTAC 71

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
Db   72 CAGGGTCGTCTGTGGGCATTCTGTTGC 98

RESULT 10
US-08-282-030-61/c
; Sequence 61, Application US/08282030
; Patent No. 5589364
; GENERAL INFORMATION:
;   APPLICANT: Williams, Jon I.
;   APPLICANT: Pierce, James C.
;   APPLICANT: Anderson, Mark G.
;   APPLICANT: Kari, Prasad
;   TITLE OF INVENTION: Recombinant Production of Biologically
;   TITLE OF INVENTION: Active Peptides and Proteins
;   NUMBER OF SEQUENCES: 62
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;   ADDRESSEE: Dunner
;   STREET: 1300 I Street, N.W.
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20005-3315
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/282,030
;   FILING DATE: 29-JUL-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Fordis, Jean B.
;   REGISTRATION NUMBER: 32,984
;   REFERENCE/DOCKET NUMBER: 05387.0001-00000
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-408-4000
;   TELEFAX: 202-408-4400
;   INFORMATION FOR SEQ ID NO: 61:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 109 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-282-030-61

Alignment Scores:
Pred. No.: 0.000738   Length: 109
Score: 80.00   Matches: 13
Percent Similarity: 55.17% Conservative: 3
Best Local Similarity: 44.83% Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: 1   Gaps: 0

US-10-045-180A-6 (1-31) x US-08-282-030-61 (1-109)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21

```
Db    102  TGTTACTCGCCGTATTCCGGCATGCATCGCAGGCGAGCGTCGCTATGGTACTTGTATTAC  43
              ||:::||||||::   ||||| |||||||
Qy     22  LeuGlyGluArgTyrProIleCysCys  30
              |||        :::   ||||||
Db     42  CAGGGTCGTCGTGGGCATTCTGTTGC  16
```

RESULT 11
PCT-US95-10219-61/c
; Sequence 61, Application PC/TUS9510219
; GENERAL INFORMATION:
";   APPLICANT: Williams, Jon I.
";   APPLICANT: Pierce, James C.
";   APPLICANT: Anderson, Mark G.
";   APPLICANT: Kari Prasad
";   TITLE OF INVENTION: Recombinant Production of Biologically
";   TITLE OF INVENTION: Active Peptides and Proteins
";   NUMBER OF SEQUENCES: 62
";   CORRESPONDENCE ADDRESS:
";   ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
";   ADDRESSEE: Dunner
";   STREET: 1300 I Street, N.W.
";   CITY: Washington
";   STATE: D.C.
";   COUNTRY: USA
";   ZIP: 20005-3315
";   COMPUTER READABLE FORM:
";   MEDIUM TYPE: Floppy disk
";   COMPUTER: IBM PC compatible
";   OPERATING SYSTEM: PC-DOS/MS-DOS
";   SOFTWARE: PatentIn Release #1.0, Version #1.25
";   CURRENT APPLICATION DATA:
";   APPLICATION NUMBER: PCT/US95/10219
";   FILING DATE:
";   CLASSIFICATION:
";   PRIOR APPLICATION DATA:
";   APPLICATION NUMBER: US 08/282,030
";   FILING DATE: 29-JUL-1994
";   ATTORNEY/AGENT INFORMATION:
";   NAME: Fordis, Jean B.
";   REGISTRATION NUMBER: 32,984
";   REFERENCE/DOCKET NUMBER: 05387.0001-00000
";   TELECOMMUNICATION INFORMATION:
";   TELEPHONE: 202-408-4000
";   TELEFAX: 202-408-4400
";   INFORMATION FOR SEQ ID NO: 61:
";   SEQUENCE CHARACTERISTICS:
";   LENGTH: 109 base pairs
";   TYPE: nucleic acid
";   STRANDEDNESS: single
";   TOPOLOGY: linear
PCT-US95-10219-61

Alignment Scores:			
Pred. No.:	0.000738	Length:	109
Score:	80.00	Matches:	13
Percent Similarity:	55.17%	Conservative:	3
Best Local Similarity:	44.83%	Mismatches:	13
Query Match:	41.88%	Indels:	0
DB:		Gaps:	5

US-10-045-180A-6 (1-31) x PCT-US95-10219-61 (1-109)

```
Qy    2  CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle  21
            ||:::||||||:::    |||||| ||||||
Db  102  TGTTACTCGCCGTATTCCGGCATGCATCGCAGGCGAGCGTCGCTATGGTACTTGTATTAC  43

Qy   22  LeuGlyGluArgTyrProIleCysCys  30
          |||        :::    ||||||
Db   42  CAGGGTCGTCGTGGGCATTCTGTTGC  16
```

RESULT 12
US-08-021-608D-5/c

; Sequence 5, Application US/08021608D
; Patent No. 5580760
; GENERAL INFORMATION:
";   APPLICANT: LEVENS, DAVID L., DUNCAN,
";   APPLICANT: ROBERT C., AND AVIGAN, MARK I.
";   TITLE OF INVENTION: NOVEL FUSE BINDING
";   TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
";   NUMBER OF SEQUENCES: 24
";   CORRESPONDENCE ADDRESS:
";   ADDRESSEE: MORGAN & FINNEGAN
";   STREET: 345 PARK AVENUE
";   CITY: NEW YORK
";   STATE: NEW YORK
";   COUNTRY: USA
";   ZIP: 10154
";   COMPUTER READABLE FORM:
";   MEDIUM TYPE: FLOPPY DISK
";   COMPUTER: IBM PC COMPATIBLE
";   OPERATING SYSTEM: PC-DOS/MS-DOS
";   SOFTWARE: WORDPERFECT 5.1
";   CURRENT APPLICATION DATA:
";   APPLICATION NUMBER: US/08/021,608D
";   FILING DATE: 22-FEB-1993
";   CLASSIFICATION: 435
";   PRIOR APPLICATION DATA:
";   APPLICATION NUMBER:
";   FILING DATE:
";   CLASSIFICATION: 435
";   ATTORNEY/AGENT INFORMATION:
";   NAME: WILLIAM S. FEILER
";   REGISTRATION NUMBER: 26,728
";   REFERENCE/DOCKET NUMBER: 2026-4063
";   TELECOMMUNICATION INFORMATION:
";   TELEPHONE: (212) 758-4800
";   TELEFAX: (212) 751-6849
";   INFORMATION FOR SEQ ID NO: 5:
";   SEQUENCE CHARACTERISTICS:
";   LENGTH: 1097
";   TYPE: Nucleic acid
";   STRANDEDNESS: Double
";   TOPOLOGY: Unknown
";   MOLECULE TYPE: cDNA
";   HYPOTHETICAL: NO
";   ANTI-SENSE: YES
";   ORIGINAL SOURCE:
";   ORGANISM: Human
";   CELL LINE: HL60
US-08-021-608D-5

Alignment Scores:			
Pred. No.:	0.0135	Length:	1097
Score:	80.00	Matches:	13
Percent Similarity:	55.17%	Conservative:	3
Best Local Similarity:	44.83%	Mismatches:	13
Query Match:	41.88%	Indels:	0
DB:	1	Gaps:	0

US-10-045-180A-6 (1-31) x US-08-021-608D-5 (1-1097)

```
Qy    2  CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle  21
            ||:::||||||:::    |||||| |||||||
Db  891  TGTTATTGCAGAATACCACGGTGCATTGCAGGAGAACGTCGCTATGGAACCTGCATCTAC  832

Qy   22  LeuGlyGluArgTyrProIleCysCys  30
          |||        :::    ||||||
Db  831  CAGGGAAGACTCTGGGCATTCTGCTGC  805
```

RESULT 13
US-08-726-160-5/c
; Sequence 5, Application US/08726160
; Patent No. 5734016
; GENERAL INFORMATION:
";   APPLICANT: LEVENS, DAVID L., DUNCAN,

; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND cDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,160
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/021,608
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1097
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: No
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
US-08-726-160-5

Alignment Scores:
Pred. No.:   0.0135   Length: 1097
Score: 80.00   Matches: 13
Percent Similarity: 55.17%   Conservative: 3
Best Local Similarity: 44.83%   Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: Gaps: 1

US-10-045-180A-6 (1-31) x US-08-726-160-5 (1-1097)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
         ||:::||||||::: |||||| |||||||
Db   891 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTATGAACCTGCATCTAC 832

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
         ||| :::
Db   831 CAGGGAAGACTCTGGGCATTCTGCTGC 805

RESULT 14
PCT-US94-01782-5/c
; Sequence 5, Application PC/TUS9401782
; GENERAL INFORMATION:
; APPLICANT: THE GOVERNMENT OF THE UNITED STATES
; APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
; APPLICANT: HEALTH AND HUMAN SERVICES
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND cDNA THEREFOR
; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01782
; FILING DATE: 22-FEB-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/021,608
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1097
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: No
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE: HL60
; ORGANELLE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
PCT-US94-01782-5

Alignment Scores:
Pred. No.:   0.0135   Length: 1097
Score: 80.00   Matches: 13
Percent Similarity: 55.17%   Conservative: 3
Best Local Similarity: 44.83%   Mismatches: 13
Query Match: 41.88%   Indels: 0
DB: 5   Gaps: 0

US-10-045-180A-6 (1-31) x PCT-US94-01782-5 (1-1097)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
         ||:::||||||::: |||||| |||||||
Db   891 TGCTATTGCAGAATACCAGCGTGCATTGCAGGAGAACGTCGCTATGAACCTGCATCTAC 832

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
         ||| :::
Db   831 CAGGGAAGACTCTGGGCATTCTGCTGC 805

RESULT 15
US-08-282-030-50
; Sequence 50, Application US/08282030
; Patent No. 5589364
GENERAL INFORMATION:
    APPLICANT: Williams, Jon I.
    APPLICANT: Pierce, James C.
    APPLICANT: Anderson, Mark G.
    APPLICANT: Kari, Prasad
    TITLE OF INVENTION: Recombinant Production of Biologically
    TITLE OF INVENTION: Active Peptides and Proteins
    NUMBER OF SEQUENCES: 62
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
        ADDRESSEE: Dunner
        STREET: 1300 I Street, N.W.
        CITY: Washington
        STATE: D.C.
        COUNTRY: USA
        ZIP: 20005-3315
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/282,030
        FILING DATE: 29-JUL-1994
        CLASSIFICATION: 435
    ATTORNEY/AGENT INFORMATION:
        NAME: Fordis, Jean B.
        REGISTRATION NUMBER: 32,984
        REFERENCE/DOCKET NUMBER: 05387.0001-00000
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 202-408-4000
        TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 50:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 123 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: single
        TOPOLOGY: linear
US-08-282-030-50

Alignment Scores:
Pred. No.:   0.00638   Length: 123
Score: 74.00   Matches: 13
Percent Similarity: 51.72%   Conservative: 2
Best Local Similarity: 44.83%   Mismatches: 14
Query Match: 38.74%   Indels: 0
DB: 1   Gaps: 0

US-10-045-180A-6 (1-31) x US-08-282-030-50 (1-123)

Qy   2 CysHisCysArgValLeuTyrCysIlePheGlyGluHisLeuGlyGlyThrCysPheIle 21
           ||::::||||| || ||| ||| |||
Db   24 TGCTACTGTGCCACTGGTCGTTGTGCAACGCGTGAAAGCCTGAGCGGCGTCTGTGAAATC 83

Qy   22 LeuGlyGluArgTyrProIleCysCys 30
           ||| || :::||||||
Db   84 TCCGGTCGTCTGTATCGCCTGTGTTGC 110

Search completed: December 17, 2003, 14:54:03
Job time : 32 secs

```
                 GenCore version 5.1.6
        Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     December 17, 2003, 14:50:14 ; Search time 20.304 Seconds
                                         (without alignments)
                                         445.225 Million cell updates/sec

Title:           US-10-045-180A-3
Perfect score:   94
Sequence:      1 MRTLTLLSAFLLVALQAWAE.........GEHLGGTCFILGERYPICCY 94

Scoring table:   OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched:        283308 seqs, 96168682 residues

Word size :      0

Total number of hits satisfying chosen parameters:        283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :       PIR_76:*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match	Length	DB	ID	Description
1	8	8.5	94	94	1	A40499	defensin alpha-1 p
2	8	8.5	94	94	1	B40499	defensin alpha-3 p
3	8	8.5		128	2	T15912	hypothetical prote
4	7	7.4		87	2	H64521	hypothetical prote
5	7	7.4		87	2	D71984	hypothetical prote
6	7	7.4		94	2	A44454	hypothetical prote
7	7	7.4		95	1	WTRBM1	macrophage antibio
8	7	7.4		95	2	WTRBM2	defensin CS-4 prec
9	7	7.4		100	1	S27016	defensin alpha-6 p
10	7	7.4		183	2	D81396	MAF homolog Cj0507
11	7	7.4		196	2	A48150	hibernation-relate
12	7	7.4		235	2	H96023	probable phosphona
13	7	7.4		286	2	A28333	carcinoembryonic a
14	7	7.4		321	2	JH0395	biliary glycoprote
15	7	7.4		328	1	S16300	UDPglucose 4-epime
16	7	7.4		328	2	A95975	probable UDPglucos
17	7	7.4		328	2	S60256	TGF-beta receptor
18	7	7.4		328	2	H84900	hypothetical prote
19	7	7.4		339	2	T34925	ABC transporter in
20	7	7.4		347	2	B71347	probable catabolit
21	7	7.4		349	2	A34815	carcinoembryonic a
22	7	7.4		351	2	JH0396	biliary glycoprote
23	7	7.4		412	2	G81581	tyrosyl-tRNA synth
24	7	7.4		412	2	H71560	biliary glycoprote
25	7	7.4		417	2	JH0394	tyrosyl tRNA synth
26	7	7.4		422	2	G86535	biliary glycoprote
27	7	7.4		422	2	D72088	tyrosyl tRNA synth
28	7	7.4		436	2	T36432	probable transmemb
29	7	7.4		464	2	C30127	transmembrane carc
30	7	7.4		526	1	A32164	biliary glycoprote
31	7	7.4		628	2	A87596	hypothetical prote
32	7	7.4		702	2	A36319	carcinoembryonic a
33	7	7.4		1450	2	JC6139	cystic fibrosis tr
34	7	7.4		1476	1	A39901	cystic fibrosis tr
35	7	7.4		1476	1	A40303	cystic fibrosis tr
36	7	7.4		1480	1	DVHUCF	cystic fibrosis tr
37	7	7.4		1485	1	S23756	CFTR protein - Afr
38	6	6.4		31	2	T09121	hypothetical prote
39	6	6.4		57	2	D86108	hypothetical prote
40	6	6.4		85	2	JC2290	hypothetical prote
41	6	6.4		85	1	AD1537	B. subtilis protei
42	6	6.4		89	2	S25602	ubiquinol-cytochro
43	6	6.4		90	2	T28989	hypothetical prote
44	6	6.4		97	2	A47365	defensin alpha-4 p
45	6	6.4		103	1	IKECV	colicin V - Escher
46	6	6.4		106	2	A88650	protein C09G12.5 [
47	6	6.4		115	2	G64304	hypothetical prote
48	6	6.4		117	2	F69887	hypothetical prote
49	6	6.4		124	2	S40099	hypothetical prote
50	6	6.4		124	2	H82117	hypothetical prote
51	6	6.4		131	2	D69827	hypothetical prote
52	6	6.4		131	2	AG3198	hypothetical prote
53	6	6.4		134	2	JC5116	interleukin-5 prec
54	6	6.4		137	2	A70913	hypothetical prote
55	6	6.4		142	2	D72340	hypothetical prote
56	6	6.4		144	2	T18016	hypothetical prote
57	6	6.4		157	2	A31084	Gdx protein - huma
58	6	6.4		161	2	F70006	cysteine dioxygena
59	6	6.4		165	2	C69065	hypothetical prote
60	6	6.4		168	2	T17592	hypothetical prote
61	6	6.4		168	2	AG2986	conserved hypothet
62	6	6.4		168	2	H98296	hypothetical prote
63	6	6.4		172	2	H81375	hypothetical prote
64	6	6.4		174	2	A49181	alpha B-crystallin
65	6	6.4		174	2	S58758	alpha-crystallin c
66	6	6.4		175	2	A70952	hypothetical prote
67	6	6.4		176	2	T21847	hypothetical prote
68	6	6.4		179	2	E64460	peptidoglycan asso
69	6	6.4		179	2	C69467	conserved hypothet
70	6	6.4		180	2	I48129	Xe169 (escapes X-i
71	6	6.4		180	2	I84489	escapes X-inactiva
72	6	6.4		181	2	AF2468	50S ribosomal prot
73	6	6.4		183	2	AF0904	probable exported
74	6	6.4		184	2	T40305	hypothetical prote
75	6	6.4		186	2	T11384	ATPase subunit 6 -
76	6	6.4		187	2	D85562	hypothetical prote
77	6	6.4		187	2	H64796	ybeM protein - Esc
78	6	6.4		189	2	T51719	hypothetical prote
79	6	6.4		191	2	F82969	hypothetical prote
80	6	6.4		195	2	G64017	hypothetical prote
81	6	6.4		195	2	D91059	hypothetical prote
82	6	6.4		195	2	B85904	hypothetical prote
83	6	6.4		197	2	A65036	hypothetical 21.2
84	6	6.4		197	2	AI0115	conserved hypothet
85	6	6.4		197	2	T26829	hypothetical prote
86	6	6.4		205	2	C88809	protein Y43C5B.1 [
87	6	6.4		207	2	C64509	hypothetical prote
88	6	6.4		212	2	T38961	hypothetical prote
89	6	6.4		213	2	AG0854	probable sugar ald
90	6	6.4		225	2	F84581	copia-like retroel
91	6	6.4		225	2	T51607	glutathione transf
92	6	6.4		226	2	C83316	NADH dehydrogenase
93	6	6.4		226	2	AC1896	hypothetical prote
94	6	6.4		226	2	D49591	membrane protein M
95	6	6.4		226	2	S37434	membrane glycoprot
96	6	6.4		226	2	C90877	probable tail asse
97	6	6.4		226	2	D90908	probable tail asse
98	6	6.4		227	2	F85842	probable tail comp
99	6	6.4		227	2	E69079	conserved hypothet
100	6	6.4		231	2	T12078	ribonuclease (EC 3

ALIGNMENTS

RESULT 1

A40499

defensin alpha-1 precursor, myeloid-related [validated] - human

N;Alternate names: defensin HP-1; myeloid leukocyte protein mrs; myeloid-related protein

N;Contains: defensin HNP-1; defensin HNP-3

C;Species: Homo sapiens (man)

C;Date: 24-Jan-1992 #sequence revision 24-Jan-1992 #text change 08-Dec-2000

C;Accession: S32499; S07161; A40499; A30658; B30658; A32958; S65412; S65414

R;Linzmeier, R.; Michaelson, D.; Liu, L.; Ganz, T.

FEBS Lett. 321, 267-273, 1993

A;Title: The structure of neutrophil defensin genes.

A;Reference number: S32499; MUID:93238968; PMID:8477861

A;Accession: S32499

A;Molecule type: DNA

A;Residues: 1-94 <LIN>

A;Cross-references: GB:L12690; NID:g292362; PIDN:AAA36382.1; PID:g292363

R;Mars, W.M.; van Tuinen, P.; Drabkin, H.A.; White, J.W.; Saunders, G.F.

Blood 71, 1713-1719, 1988

A;Title: A myeloid-related sequence that localizes to human chromosome 8q21.1-22.

A;Reference number: S07161; MUID:88222410; PMID:3370315

A;Accession: S07161

A;Molecule type: mRNA

A;Residues: 1-94 <MAR>

A;Cross-references: GB:M26602; NID:g181528; PIDN:AAA52303.1; PID:g181529

R;Daher, K.A.; Lehrer, R.I.; Ganz, T.; Kronenberg, M.

Proc. Natl. Acad. Sci. U.S.A. 85, 7327-7331, 1988

A;Title: Isolation and characterization of human defensin cDNA clones.

A;Reference number: A40499; MUID:89017191; PMID:3174637

A;Accession: A40499

A;Molecule type: mRNA

A;Residues: 1-94 <DAH>

A;Cross-references: GB:M21130; GB:M22160; NID:g181526; PIDN:AAA52302.1; PID:g181527

R;Selsted, M.E.; Harwig, S.S.L.; Ganz, T.; Schilling, J.W.; Lehrer, R.I.

J. Clin. Invest. 76, 1436-1439, 1985

A;Title: Primary structures of three human neutrophil defensins.

A;Reference number: A30658; MUID:86034582; PMID:4056036

A;Accession: A30658

A;Molecule type: protein

A;Residues: 65-94 <SEL1>

A;Experimental source: HNP-1

A;Accession: B30658

A;Molecule type: protein

A;Residues: 66-94 <SEL2>

A;Experimental source: HNP-2

R;Selsted, M.E.; Harwig, S.S.L.

J. Biol. Chem. 264, 4003-4007, 1989

A;Title: Determination of the disulfide array in the human defensin HNP-2. A covalently

A;Reference number: A32958; MUID:89139465; PMID:2917986

A;Accession: A32958

A;Molecule type: protein

A;Residues: 66-94 <SEL>

R;Frohm, M.; Gunne, H.; Bergman, A.C.; Agerberth, B.; Bergman, T.; Boman, A.; Liden, S.;

Eur. J. Biochem. 237, 86-92, 1996

A;Title: Biochemical and antibacterial analysis of human wound and blister fluid.

A;Reference number: S65409; MUID:96203912; PMID:8620898

A;Accession: S65412

A;Molecule type: protein

A;Residues: 65,'X',67,'X',69-72,'X',74-82 <FRO1>

A;Accession: S65414

A;Molecule type: protein

A;Residues: 'X',67,'X',69-72,'X',74-82,84 <FRO2>

R;Pardi, A.; Zhang, X.L.; Selsted, M.E.; Skalicky, J.J.; Yip, P.F.

Biochemistry 31, 11357-11364, 1992

A;Title: NMR studies of defensin antimicrobial peptides. 2. Three-dimensional structures

A;Reference number: A44209; MUID:93075734; PMID:1445873

A;Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR

R;Wilson, C.L.; Ouellette, A.J.; Satchell, D.P.; Ayabe, T.; Lopez-Boado, Y.S.; Stratman,

Science 286, 113-117, 1999

A;Title: Regulation of intestinal alpha-defensin activation by the metalloproteinase mat

A;Reference number: A59079; MUID:99438247; PMID:10506557

A;Contents: annotation; activation by matrilysin

C;Comment: The propeptide is cleaved by matrilysin (EC 3.4.24.23) (see PIR:KCHUM).

C;Genetics:

A;Gene: GDB:DEFA1; DEF1; MRS; HNP-1

A;Cross-references: GDB:120745; OMIM:125220

A;Map position: 8p23-8p22

A;Introns: 59/1

C;Complex: homodimer

C;Superfamily: mammalian defensin

C;Keywords: antibiotic; homodimer

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-64/Domain: propeptide #status predicted <PRO>

F;65-94/Product: defensin HNP-1 #status experimental <MAT1>

F;66-94/Product: defensin HNP-2 #status experimental <MAT2>

F;66-94,68-83,73-93/Disulfide bonds: #status experimental

Query Match   8.5%; Score 8; DB 1; Length 94;

Best Local Similarity 100.0%; Pred. No. 0.53;

Matches   8; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   19 AEPLQARA 26
          ||||||||
Db   19 AEPLQARA 26

RESULT 2

B40499

defensin alpha-3 precursor, neutrophil-specific [validated] - human

N;Alternate names: defensin HNP-3; neutrophil peptide 3 (HNP-3)

N;Contains: defensin HNP-2; defensin HNP-3

C;Species: Homo sapiens (man)

C;Date: 24-Jan-1992 #sequence revision 03-Nov-1995 #text change 15-Sep-2000

C;Accession: S32500; S08005; B40499; S65413; A32958; C30658

R;Linzmeier, R.; Michaelson, D.; Liu, L.; Ganz, T.

FEBS Lett. 321, 267-273, 1993

A;Title: The structure of neutrophil defensin genes.

A;Reference number: S32499; MUID:93238968; PMID:8477861

A;Accession: S32500

A;Molecule type: DNA

A;Residues: 1-94 <LIN>

A;Cross-references: GB:L12691; NID:g292364; PIDN:AAB57722.1; PID:g292365

R;Wiedemann, L.M.; Francis, G.E.; Lamb, R.F.; Burns, J.H.; Winnie, J.N.; MacKenzie, E.D

submitted to the EMBL Data Library, November 1988

A;Reference number: S08005

A;Accession: S08005

A;Molecule type: mRNA

A;Residues: 1-94 <WIE>

A;Cross-references: EMBL:X13621; NID:g30500; PIDN:CAA31952.1; PID:g30501

A;Experimental source: HNP-3 expressed in promyelocytes and myelocytes

R;Daher, K.A.; Lehrer, R.I.; Ganz, T.; Kronenberg, M.

Proc. Natl. Acad. Sci. U.S.A. 85, 7327-7331, 1988

A;Title: Isolation and characterization of human defensin cDNA clones.

A;Reference number: A40499; MUID:89017191; PMID:3174637

A;Accession: B40499

A;Molecule type: mRNA

A;Residues: 30-94 <DAH>

A;Cross-references: GB:M21131; NID:g181530; GB:M23281; NID:g181534; PIDN:AAA52304.1; PI

A;Experimental source: clone HNP-3

A;Note: the sequence in GenBank entry HUMDEF3A, release 111.0, is translated in an inco

A;Note: the sequence in GenBank entry HUMDEFB, release 111.0, has been corrected to cor

R;Frohm, M.; Gunne, H.; Bergman, A.C.; Agerberth, B.; Bergman, T.; Boman, A.; Liden, S.

Eur. J. Biochem. 237, 86-92, 1996

A;Title: Biochemical and antibacterial analysis of human wound and blister fluid.

A;Reference number: S65409; MUID:96203912; PMID:8620898

A;Accession: S65413

A;Molecule type: protein

A;Residues: 65,'X',67,'X',69-72,'X',74-82 <FRO>

R;Selsted, M.E.; Harwig, S.S.L.

J. Biol. Chem. 264, 4003-4007, 1989

A;Title: Determination of the disulfide array in the human defensin HNP-2. A covalently

A;Reference number: A32958; MUID:89139465; PMID:2917986

A;Accession: A32958

A;Molecule type: protein

```
                    GenCore version 5.1.6

         Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - protein search, using sw model

Run on:   December 17, 2003, 14:50:14 ; Search time 15.792 Seconds
                               (without alignments)
                          279.921 Million cell updates/sec

Title:          US-10-045-180A-3
Perfect score:  94
Sequence:       1 MRTLTLLSAFLLVALQAWAE.........GEHLGGTCFILGERYPICCY 94

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0
Searched:       127863 seqs, 47026705 residues
Word size :     0
Total number of hits satisfying chosen parameters:      127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 100 summaries
Database :  SwissProt_41:*

          Pred. No. is the number of results predicted by chance to have a
          score greater than or equal to the score of the result being printed,
          and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Match	Query Match Length	DB Length	ID	Description
1	8.5	8	94	1	DEF1_HUMAN	P59665 homo sapien
2	8.5	8	94	1	DEF3_HUMAN	P59666 homo sapien
3	8.5	8	649	1	S281_HUMAN	O00337 homo sapien
4	7.4	7	87	1	DEFA_RAT	Q62713 rattus norv
5	7.4	7	93	1	DEF4_RAT	Q62714 rattus norv
6	7.4	7	94	1	DEF1_RAT	Q62716 rattus norv
7	7.4	7	94	1	DEF2_RAT	Q62715 rattus norv
8	7.4	7	94	1	DEF5_HUMAN	Q01523 homo sapien
9	7.4	7	95	1	DEF3_RABIT	P01376 oryctolagus
10	7.4	7	95	1	DEF4_RABIT	P01377 oryctolagus
11	7.4	7	100	1	DEF6_HUMAN	Q01524 homo sapien
12	7.4	7	183	1	PSAF_PROMA	Q9x714 prochloroco
13	7.4	7	183	1	Y507_CAMJE	Q9pi04 campylobact
14	7.4	7	196	1	HP20_TAMSI	Q06575 tamias sibi
15	7.4	7	327	1	EXOB_RHILT	Q59745 rhizobium l
16	7.4	7	328	1	EXOB_RHIME	P26503 rhizobium m
17	7.4	7	328	1	IF32_ARATH	Q38884 arabidopsis
18	7.4	7	349	1	CEA8_HUMAN	P31997 homo sapien
19	7.4	7	412	1	SYY_CHLPN	Q9z812 chlamydia p
20	7.4	7	412	1	SYY_CHLTR	O84065 chlamydia t
21	7.4	7	526	1	CEA1_HUMAN	P13688 homo sapien
22	7.4	7	702	1	CEA5_HUMAN	P06731 homo sapien
23	7.4	7	1450	1	CFTR_RABIT	Q00554 oryctolagus
24	7.4	7	1476	1	CFTR_MOUSE	P26361 mus musculu
25	7.4	7	1480	1	CFTR_HUMAN	P13569 homo sapien
26	7.4	7	1485	1	CFTR_XENLA	P26363 xenopus lae
27	6.4	6	31	1	PETL_NEPOL	Q9tky9 nephroselmi
28	6.4	6	31	1	PETL_POPDE	O20272 populus del
29	6.4	6	52	1	CRAB_COLLI	O12987 columba liv
30	6.4	6	52	1	CRAB_TURME	O12995 turdus meru
31	6.4	6	57	1	YDO_ECOLI	P58038 escherichia
32	6.4	6	67	1	Y12B_BPT4	P39494 bacteriopha
33	6.4	6	85	1	Y522_PSEAE	Q51483 pseudomonas
34	6.4	6	97	1	DEF4_HUMAN	P12838 homo sapien
35	6.4	6	103	1	CBAV_ECOLI	P22522 escherichia
36	6.4	6	115	1	Y039_METJA	Q60351 methanococc
37	6.4	6	131	1	CRB2_BACSU	O07591 bacillus su
38	6.4	6	132	1	IL5_SHEEP	Q28586 ovis aries
39	6.4	6	134	1	IL5_BOVIN	P52173 bos taurus
40	6.4	6	134	1	IL5_CERTO	P46685 cercocebus
41	6.4	6	134	1	IL5_HORSE	O02699 equus cabal
42	6.4	6	135	1	IL5_CAVPO	O08987 cavia porce
43	6.4	6	137	1	Y049_MYCTU	P71706 mycobacteri
44	6.4	6	153	1	VNSC_TPMV	Q9ws38 tupaia para
45	6.4	6	157	1	UBLG_HUMAN	P11441 homo sapien
46	6.4	6	174	1	CRAB_ANAPL	Q05557 anas platyr
47	6.4	6	174	1	CRAB_CHICK	Q05713 gallus gall
48	6.4	6	180	1	SMCX_CRIGR	P41228 cricetulus
49	6.4	6	181	1	BY55_HUMAN	O95971 homo sapien
50	6.4	6	181	1	RL10_ANASP	Q8y116 anabaena sp
51	6.4	6	185	1	BY55_MOUSE	O88875 mus musculu
52	6.4	6	191	1	Y977_HAEIN	P44088 haemophilus
53	6.4	6	195	1	YFIK_ECOLI	P38101 escherichia
54	6.4	6	197	1	HAM1_YERPE	Q8zhf4 yersinia pe
55	6.4	6	205	1	YG77_METJA	Q59071 methanococc
56	6.4	6	207	1	SEPB_SCHPO	O14198 schizosacch
57	6.4	6	210	1	PSBP_PORPU	P50272 porphyra pu
58	6.4	6	233	1	YUT5_YEAST	P39541 saccharomyc
59	6.4	6	253	1	EBSD_ENTFA	P36923 enterococcu
60	6.4	6	254	1	AROD_NEIMA	Q9jtr9 neisseria m
61	6.4	6	254	1	AROD_NEIMB	Q9jyt0 neisseria m
62	6.4	6	258	1	CTR2_ANOGA	Q17025 anopheles g
63	6.4	6	259	1	CTR1_ANOGA	Q27289 anopheles g
64	6.4	6	261	1	COX3_PAPHA	Q9zxx8 papio hamad
65	6.4	6	261	1	VGL1_CVPR8	P33464 porcine res
66	6.4	6	262	1	VME1_CVCAI	P36299 canine ente
67	6.4	6	262	1	VME1_CVPFS	P09175 porcine tra
68	6.4	6	262	1	VME1_CVPPU	P04135 porcine tra
69	6.4	6	262	1	VME1_CVPRM	P24412 porcine res
70	6.4	6	262	1	VME1_FIPV	P25878 feline infe
71	6.4	6	262	1	YBEM_ECO57	P58054 escherichia
72	6.4	6	262	1	YBEM_ECOLI	P39874 escherichia
73	6.4	6	269	1	COX3_MAGGR	Q95840 magnaporthe
74	6.4	6	277	1	CYST_ECOLI	P16701 homo sapien
75	6.4	6	291	1	SH1D_PIG	P79400 sus scrofa
76	6.4	6	294	1	BLAC_CITDI	P22390 citrobacter
77	6.4	6	294	1	MC5R_PIG	Q9mzv8 sus scrofa
78	6.4	6	296	1	YFCH_HAEIN	P71373 haemophilus
79	6.4	6	302	1	DAPA_XYLFA	Q9per5 xylella fas
80	6.4	6	303	1	T2RD_HUMAN	Q9nyv9 homo sapien
81	6.4	6	306	1	YC64_AQUAE	O67303 aquifex aeo
82	6.4	6	309	1	TR43_HUMAN	P59537 homo sapien
83	6.4	6	309	1	TR52_HUMAN	P59546 homo sapien
84	6.4	6	316	1	T2BB_BACSU	P33562 bacillus su
85	6.4	6	332	1	Y4RC_RHISN	P55636 rhizobium s
86	6.4	6	336	1	XYLZ_PSEPU	P23101 pseudomonas
87	6.4	6	338	1	Y941_METJA	Q57711 methanococc
88	6.4	6	342	1	ISIA_SYNP2	P31157 synechococc
89	6.4	6	344	1	CEA6_HUMAN	P40199 homo sapien
90	6.4	6	346	1	DHAS_BACSU	Q04797 bacillus su
91	6.4	6	347	1	NU2M_DIDMA	P41305 didelphis m
92	6.4	6	349	1	SOHB_ECOLI	P24213 escherichia
93	6.4	6	356	1	LPXB_AQUAE	O67420 aquifex aeo
94	6.4	6	362	1	NTCP_MOUSE	O08705 mus musculu
95	6.4	6	364	1	PLCE_HUMAN	Q9nug2 homo sapien
96	6.4	6	365	1	FUT3_BOVIN	Q11126 bos taurus
97	6.4	6	378	1	FETB_RAT	Q9qx79 rattus norv
98	6.4	6	380	1	CYB_PONPA	P92701 pongo pygma
99	6.4	6	380	1	CYB_PONPP	P92723 pongo pygma
100	6.4	6	380	1	CYB_PONPY	Q35614 pongo pygma

```
                              ALIGNMENTS

RESULT 1
```

```
ID   DEF1_HUMAN     STANDARD;      PRT;     94 AA.
AC   P59665; P11479; Q14125;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Neutrophil defensin 1 precursor (HNP-1) (HP1) (Defensin, alpha
DE   1) [Contains: HP 1-56; Neutrophil defensin 2 (HNP-2) (HP2)].
GN   DEFA1 OR DEF1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89017191; PubMed=3174637;
RA   Daher K.A., Lehrer R.I., Ganz T., Kronenber M.;
RT   "Isolation and characterization of human defensin cDNA clones.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:7327-7331(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88222410; PubMed=3370315;
RA   Mars W.M., Vantuinen P., Drabkin H.A., White J.W., Saunders G.F.;
RT   "A myeloid-related sequence that localizes to human chromosome
RT   8q21.1-22.";
RL   Blood 71:1713-1719(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89142618; PubMed=2918759;
RA   Wiedemann L.M., Francis G.E., Lamb R.F., Burns J.H., Winnie J.N.,
RA   McKenzie E.D., Birnie G.D.;
RT   "Differentiation stage-specific expression of a gene during
RT   granulopoiesis.";
RL   Leukemia 3:227-234(1989).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96107186; PubMed=8530462;
RA   Mars W.M., Patmasiriwat P., Maity T., Huff V., Weil M.M.,
RA   Saunders G.F.;
RT   "Inheritance of unequal numbers of the genes encoding the human
RT   neutrophil defensins HP-1 and HP-3.";
RL   J. Biol. Chem. 270:30371-30376(1995).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93238968; PubMed=8477861;
RA   Linzmeier R., Michaelson D., Liu L., Ganz T.;
RT   "The structure of neutrophil defensin genes.";
RL   FEBS Lett. 321:267-273(1993).
RN   [6]
RP   ERRATUM.
RX   MEDLINE=93314809; PubMed=8325384;
RA   Linzmeier R., Michaelson D., Liu L., Ganz T.;
RL   FEBS Lett. 326:299-300(1993).
RN   [7]
RP   SEQUENCE OF 65-94.
RX   MEDLINE=86034582; PubMed=4056036;
RA   Selsted M.E., Harwig S.S.L., Ganz T., Schilling J.W., Lehrer R.I.;
RT   "Primary structures of three human neutrophil defensins.";
RL   J. Clin. Invest. 76:1436-1439(1985).
RN   [8]
RP   DISULFIDE BONDS.
RX   MEDLINE=89139465; PubMed=2917986;
RA   Selsted M.E., Harwig S.S.L.;
RT   "Determination of the disulfide array in the human defensin HNP-2. A
RT   covalently cyclized peptide.";
RL   J. Biol. Chem. 264:4003-4007(1989).
RN   [9]
RP   PROCESSING.
RX   MEDLINE=91210262; PubMed=2019582;
RA   Bateman A., Singh A., Shustik C., Mars W.M., Solomon S.;
RT   "The isolation and identification of multiple forms of the neutrophil
RT   granule peptides from human leukemic cells.";
RL   J. Biol. Chem. 266:7524-7530(1991).
RN   [10]
RP   PROCESSING.
RX   MEDLINE=92190492; PubMed=1339298;
RA   Valore E.V., Ganz T.;
RT   "Posttranslational processing of defensins in immature human myeloid
RT   cells.";
RL   Blood 79:1538-1544(1992).
RN   [11]
RP   STRUCTURE BY NMR OF DEFENSIN 1.
RX   MEDLINE=93075733; PubMed=1445872;
RA   Zhang X.-L., Selsted M.E., Pardi A.;
RT   "NMR studies of defensin antimicrobial peptides. 1. Resonance
RT   assignment and secondary structure determination of rabbit NP-2 and
RT   human HNP-1.";
RL   Biochemistry 31:11348-11356(1992).
RN   [12]
RP   STRUCTURE BY NMR OF DEFENSIN 1.
RX   MEDLINE=93075734; PubMed=1445873;
RA   Pardi A., Zhang X.-L., Selsted M.E., Skalicky J.J., Yip P.F.;
RT   "NMR studies of defensin antimicrobial peptides. 2. Three-dimensional
RT   structures of rabbit NP-2 and human HNP-1.";
RL   Biochemistry 31:11357-11364(1992).
CC   -!- FUNCTION: Defensins 1/2 have antibiotic, fungicide and antiviral
CC       activities.
CC   -!- SUBUNIT: DIMER.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M21130; AAA52302.1; -.
DR   EMBL; M26602; AAA52303.1; -.
DR   EMBL; L12690; AAA36382.1; -.
DR   EMBL; X52053; CAA36280.1; -.
DR   Genew; HGNC:2761; DEFA1.
DR   MIM; 125220; -.
DR   GO; GO:0006935; P:chemotaxis; TAS.
DR   GO; GO:0006955; P:immune response; TAS.
DR   InterPro; IPR006081; Defensin_alpha.
DR   InterPro; IPR006080; Defensin_mammal.
DR   InterPro; IPR002366; Defensin_propep.
DR   Pfam; PF00879; Defensin_propep; 1.
DR   Pfam; PF00323; defensins; 1.
DR   SMART; SM00048; DEFSN; 1.
DR   PROSITE; PS00269; DEFENSIN; 1.
KW   Defensin; Antibiotic; Antiviral; Fungicide; Signal.
FT   SIGNAL        1     19
FT   PROPEP       20     38
FT   CHAIN        39     94       HP 1-56.
FT   CHAIN        65     94       NEUTROPHIL DEFENSIN 1.
FT   CHAIN        66     94       NEUTROPHIL DEFENSIN 2.
FT   DISULFID     66     94
FT   DISULFID     68     83
FT   DISULFID     73     93
SQ   SEQUENCE   94 AA;  10201 MW;  0E0F8E95737396FA CRC64;
```

```
Query Match             8.5%;  Score 8;  DB 1;  Length 94;
Best Local Similarity 100.0%;  Pred. No. 0.31;
Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     19 AEPIQARA 26
          ||||||||
Db     19 AEPIQARA 26


RESULT 2
DEF3_HUMAN
```

```
                    GenCore version 5.1.6
         Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 17, 2003, 14:50:14 ; Search time 44.368 seconds
                                              (without alignments)
                                              546.722 Million cell updates/sec

Title:          US-10-045-180A-3
Perfect score:  94
Sequence:       1 MRTILLSAFLIVALQAWAE.........GEHLGGTCFILGERYPICCY 94

Scoring table:  OLIGO
                .Gapop 60.0 , Gapext 60.0

Searched:       830525 seqs, 258052604 residues

Word size :     0

Total number of hits satisfying chosen parameters:        830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :      SPTREMBL_23:*
         1:     sp_archea:*
         2:     sp_bacteria:*
         3:     sp_fungi:*
         4:     sp_human:*
         5:     sp_invertebrate:*
         6:     sp_mammal:*
         7:     sp_mhc:*
         8:     sp_organelle:*
         9:     sp_phage:*
        10:     sp_plant:*
        11:     sp_rodent:*
        12:     sp_virus:*
        13:     sp_vertebrate:*
        14:     sp_unclassified:*
        15:     sp_rvirus:*
        16:     sp_bacteriap:*
        17:     sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	Query Match	Length	DB	ID	Description
1	8	8.5	128	5	Q19064	Q19064 caenorhabdi
2	8	8.5	467	10	Q9FF99	Q9ff99 arabidopsis
3	8	8.5	649	4	Q96PL7	Q96pl7 homo sapien
4	7	7.4	87	16	O24861	O24861 helicobacte
5	7	7.4	87	16	Q9ZN45	Q9zn45 helicobacte
6	7	7.4	96	6	Q9TU00	Q9tu00 macaca mula
7	7	7.4	96	6	Q9TT29	Q9ttz9 macaca mula
8	7	7.4	96	16	P82318	P82318 macaca mula
9	7	7.4	111	16	Q8EJG5	Q8ejg5 shewanella
10	7	7.4	151	11	Q99KZ7	Q99kz7 mus musculu
11	7	7.4	164	16	Q98MM4	Q98mm4 rhizobium l
12	7	7.4	196	11	Q920N0	Q920n0 tamias sibi
13	7	7.4	200	2	Q9KGW7	Q9kgw7 pseudomonas
14	7	7.4	222	12	Q9QTD0	Q9qtd0 marek's dis
15	7	7.4	225	10	Q8RU05	Q8ru05 oryza sativ
16	7	7.4	235	16	Q92TQ1	Q92tq1 rhizobium m
17	7	7.4	248	10	Q8LLZ3	Q8ll23 oryza sativ
18	7	7.4	265	16	Q92RV2	Q92rv2 rhizobium m
19	7	7.4	287	4	Q13984	Q13984 homo sapien
20	7	7.4	296	16	Q8PHM1	Q8phm1 xanthomonas
21	7	7.4	318	16	Q8R9J7	Q8r9j7 thermoanaer
22	7	7.4	328	10	Q9C5Z0	Q9c520 arabidopsis
23	7	7.4	328	10	Q94K09	Q94k09 arabidopsis
24	7	7.4	339	16	O69943	O69943 streptomyce
25	7	7.4	347	16	O83285	O83285 treponema p
26	7	7.4	348	11	Q8C0P2	Q8c0p2 mus musculu
27	7	7.4	364	4	Q8NBA5	Q8nba5 homo sapien
28	7	7.4	423	4	O00717	O00717 homo sapien
29	7	7.4	430	4	Q15600	Q15600 homo sapien
30	7	7.4	436	16	Q9XAA6	Q9xaa6 streptomyce
31	7	7.4	438	10	Q8LMD1	Q8lmd1 oryza sativ
32	7	7.4	461	4	O60430	O60430 homo sapien
33	7	7.4	461	4	Q13854	Q13854 homo sapien
34	7	7.4	464	4	Q16170	Q16170 homo sapien
35	7	7.4	468	4	Q96CA7	Q96ca7 homo sapien
36	7	7.4	484	4	Q9UMN4	Q9umn4 homo sapien
37	7	7.4	547	10	Q8S5D3	Q8s5d3 oryza sativ
38	7	7.4	582	16	Q94HB7	Q94hb7 oryza sativ
39	7	7.4	628	16	Q9A4M9	Q9a4m9 caulobacter
40	7	7.4	691	10	Q943H2	Q943h2 oryza sativ
41	7	7.4	702	4	Q8N4D0	Q8n4d0 homo sapien
42	7	7.4	712	6	Q8WMU5	Q8wmus canis famil
43	7	7.4	714	10	Q8W5I4	Q8w5i4 oryza sativ
44	7	7.4	715	10	Q8S5G0	Q8s5g0 oryza sativ
45	7	7.4	765	10	Q8LHK4	Q8lhk4 oryza sativ
46	7	7.4	901	5	Q9VCX7	Q9vcx7 drosophila
47	7	7.4	948	5	Q95RT4	Q95rt4 drosophila
48	7	7.4	1242	5	Q9GRQ5	Q9grq5 leishmania
49	7	7.4	1481	6	O62673	O62673 macaca mula
50	7	7.4	1481	6	Q9TTX9	Q9ttx9 oryctolagus
51	7	7.4	1481	6	Q9TUQ2	Q9tuq2 macaca neme
52	7	7.4	1481	6	Q9TSP5	Q9tsp5 papio anubi
53	7	7.4	1481	6	O62668	O62668 macaca mula
54	7	7.4	1485	13	P70034	P70034 xenopus lae
55	7	7.4	1949	12	Q8UZB6	Q8uzb6 grapevine f
56	7	7.4	3703	5	Q8IMW2	Q8imw2 drosophila
57	7	7.4	3912	5	Q9VC56	Q9vc56 drosophila
58	6	6.4	36	8	Q959Y6	Q959y6 gymnothorax
59	6	6.4	39	13	Q8JJ30	Q8jj30 ficedula al
60	6	6.4	39	13	Q8JJ31	Q8jj31 ficedula hy
61	6	6.4	61	16	Q9RJ33	Q9rj33 streptomyce
62	6	6.4	70	16	Q8Y1Z1	Q8y1z1 ralstonia s
63	6	6.4	71	16	Q982X6	Q982x6 rhizobium l
64	6	6.4	72	12	Q8JMX9	Q8jmx9 human astro
65	6	6.4	72	16	Q8G1D2	Q8g1d2 brucella su
66	6	6.4	73	12	Q8QP26	Q8qp26 camelpox vi
67	6	6.4	77	16	Q8DAF1	Q8daf1 vibrio vuln
68	6	6.4	82	10	Q9SDE3	Q9sde3 oryza sativ
69	6	6.4	85	16	Q92DH7	Q92dh7 listeria in
70	6	6.4	86	10	Q944F0	Q944f0 oryza sativ
71	6	6.4	89	8	Q31693	Q31693 artemia par
72	6	6.4	90	8	P91520	P91520 caenorhabdi
73	6	6.4	93	16	Q8NLU4	Q8nlu4 corynebacte
74	6	6.4	95	2	Q44487	Q44487 a mo-nitrog
75	6	6.4	95	5	Q9G9T9	Q9g9t9 nymphon gra
76	6	6.4	103	4	O75215	O75215 homo sapien
77	6	6.4	106	5	O44457	O44457 caenorhabdi
78	6	6.4	107	6	Q9BG30	Q9bg30 lemur catta
79	6	6.4	107	16	Q9L9J6	Q91j6 salmonella
80	6	6.4	111	11	Q91X84	Q91x84 mus musculu
81	6	6.4	117	16	P94484	P94484 bacillus su
82	6	6.4	118	6	Q9TV10	Q9tv10 canis famil
83	6	6.4	120	8	Q8SGA1	Q8sga1 anthopleura
84	6	6.4	123	12	Q9QAJ3	Q9qaj3 equine arte
85	6	6.4	123	12	Q9QAL4	Q9qal4 equine arte
86	6	6.4	123	12	Q9QAK8	Q9qak8 equine arte
87	6	6.4	123	12	Q9QAL2	Q9qal2 equine arte
88	6	6.4	124	2	Q45924	Q45924 clostridium
89	6	6.4	124	16	Q9KQ95	Q9kq95 vibrio chol

```
90     6    6.4    125    2    Q47664              OC    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
91     6    6.4    125    2    Q8GNE4              OC    eurosids II; Brassicales; Brassicaceae; Arabidopsis.
92     6    6.4    125    16   Q8RHD8              OX    NCBI_TaxID=3702;
93     6    6.4    127    16   Q8DUL1              RN    [1]
94     6    6.4    131    16   Q8UK07              RP    SEQUENCE FROM N.A.
95     6    6.4    134    6    Q95J76              RC    STRAIN=Columbia;
96     6    6.4    134    6    Q9MYM5              RX    MEDLINE=97471969; PubMed=9330910;
97     6    6.4    134    10   Q42269              RA    Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
98     6    6.4    136    2    O69489              RA    Miyajima N., Tabata S.;
99     6    6.4    136    11   Q8C344              RT    "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
100    6    6.4    142    16   Q9WZK4              RT    features of the 1.6 Mb regions covered by twenty physically assigned
                                                   RT    P1 clones.";
                                                   RL    DNA Res. 4:215-230(1997).
                   Q47664   escherichia           DR    EMBL; AB005244; BAB10054.1; -.
                   Q8gne4   sulfurospir           DR    InterPro; IPR002422; AA/rel_permease2.
                   Q8rhd8   fusobacteri           DR    InterPro; IPR000504; RNA_rec_mot.
                   Q8dul1   streptococc           DR    Pfam; PF01490; Aa_trans; 1.
                   Q8uk07   agrobacteri           DR    PROSITE; PS00030; RRM_RNP_1; 1.
                   Q95J76   canis famil           SQ    SEQUENCE   467 AA;  51722 MW;  1B841C85CE364406 CRC64;
                   Q9mym5   sus scrofa
                   Q42269   arabidopsis           Query Match             8.5%;  Score 8;  DB 10;  Length 467;
                   O69489   mycobacteri           Best Local Similarity  100.0%;  Pred. No. 7.5;
                   Q8C344   mus musculu           Matches     8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
                   Q9wzk4   thermotoga
                                                   Qy          5 TLLSAPLL 12
                                                                 ||||||||
                    ALIGNMENTS                     Db         72 TLLSAPLL 79

RESULT 1                                           RESULT 3
Q19064                                             Q96PL7
ID   Q19064;                PRELIMINARY;    PRT;    128 AA.      ID   Q96PL7;                PRELIMINARY;    PRT;    649 AA.
AC   Q19064;                                       AC   Q96PL7;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)          DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)    DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)   DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 13.7 kDa protein.                DE   Concentrative nucleoside transporter 1.
GN   E04F6.9.                                      OS   Homo sapiens (Human).
OS   Caenorhabditis elegans.                       OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;    OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC   Rhabditidae; Peloderinae; Caenorhabditis.     OX   NCBI_TaxID=9606;
OX   NCBI_TaxID=6239;                              RN   [1]
RN   [1]                                           RP   SEQUENCE FROM N.A.
RP   SEQUENCE FROM N.A.                            RC   TISSUE=Liver;
RC   STRAIN=Bristol N2;                            RA   Mata J.F., Garcia-Manteiga J.M., Lostao P., Fernandez-Veledo S.,
RX   MEDLINE=99069613; PubMed=9851916;             RA   Larrayoz I.M., Lloberas J., Casado J., Pastor-Anglada M.;
RA   None;                                         RT   "Role of the human concentrative nucleoside transporter (hCNT1) in the
RT   "Genome sequence of the nematode C. elegans: a platform for    RT   cytotoxic action of 5'-deoxy-5-fluorouridine, an active intermediate
RT   investigating biology. The C. elegans Sequencing Consortium.";    RT   metabolite of capecitabine, a novel oral fluoropyrimidine used in
RL   Science 282:2012-2018(1998).                  RT   cancer treatment.";
RN   [2]                                           RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RP   SEQUENCE FROM N.A.                            DR   EMBL; AF309632; AAL09447.1; -.
RC   STRAIN=Bristol N2;                            DR   InterPro; IPR002668; Nuclsd_transpt2.
RA   Pauley A.;                                    DR   Pfam; PF01773; Nucleoside_tra2; 1.
RT   "The sequence of C. elegans cosmid E04F6.";   DR   ProDom; PD003768; Nuclsd_transpt2; 1.
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.    DR   TIGRFAMs; TIGR00804; nup; 1.
RN   [3]                                           DR   TIGRFAMs; TIGR00804; Nuclsd_transpt; 1.
RP   SEQUENCE FROM N.A.                            SQ   SEQUENCE   649 AA;  71598 MW;  DCD4EA8F585A00CD CRC64;
RC   STRAIN=Bristol N2;
RA   Waterston R.;                                 Query Match             8.5%;  Score 8;  DB 4;  Length 649;
RT   "Direct Submission.";                         Best Local Similarity  100.0%;  Pred. No. 9.9;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.    Matches     8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
DR   EMBL; U28943; AAA68364.1; -.
DR   WormPep; E04F6.9; CE01221.                    Qy          7 LSAFLLVA 14
KW   Hypothetical protein.                                       ||||||||
SQ   SEQUENCE   128 AA;  13694 MW;  7BBAF31DB567ACEB CRC64;    Db         94 LSAFLLVA 101

Query Match             8.5%;  Score 8;  DB 5;  Length 128;    RESULT 4
Best Local Similarity  100.0%;  Pred. No. 2.5;    O24861
Matches     8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;    ID   O24861;                PRELIMINARY;    PRT;    87 AA.
                                                   AC   O24861;
Qy          7 LSAFLLVA 14                          DT   01-JAN-1998 (TrEMBLrel. 05, Created)
              ||||||||                             DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
Db          6 LSAFLLVA 13

RESULT 2
Q9FF99
ID   Q9FF99;                PRELIMINARY;    PRT;    467 AA.
AC   Q9FF99;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Amino acid transporter.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
              GenCore version 5.1.6
         Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - protein search, using sw model

Run on:    December 17, 2003, 14:50:14 ; Search time 55.648 Seconds
                                   (without alignments)
                                   268.119 Million cell updates/sec

Title:          US-10-045-180A-3
Perfect score:  94
Sequence:       1 MRTLTLLSAFLLVALQAWAE........GEHLGGTCFILGERYPICCY 94

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       1107863 seqs, 158726573 residues

Word size :     0

Total number of hits satisfying chosen parameters:      1107863

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing: Listing first 100 summaries

Database :      A Geneseq 19Jun03:*
            1_  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
            2:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
            3:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
            4:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
            5:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
            6:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
            7:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
            8:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
            9:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
           10:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
           11:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
           12:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
           13:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
           14:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
           15:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
           16:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
           17:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
           18:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
           19:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
           20:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
           21:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
           22:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
           23:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
           24:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	94	100.0	94	20	AAY01603	Amino acid sequenc
2	94	100.0	94	20	AAY01604	Amino acid sequenc
3	31	33.0	31	20	AAY01606	Fragment of human
4	8	8.5	58	22	ABB44491	Peptide #11997 enc
5	8	8.5	58	22	AAU32095	Novel human secret
6	8	8.5	58	22	AAM78244	Human bone marrow
7	8	8.5	58	23	ABG47256	Human peptide retro
8	8	8.5	59	24	AAE33811	Human mature retro
9	8	8.5	72	19	AAW71692	Human defensin pre

No.				ID	Description
10	8	8.5	94	ABB98494	Human defensin 1.
11	8	8.5	101	ABB97309	Novel human protei
12	8	8.5	111	AAG45731	Arabidopsis thalia
13	8	8.5	118	AAG17608	Arabidopsis thalia
14	8	8.5	120	AAR15222	Chronic myelogenou
15	8	8.5	124	AAW71693	R21-Hpro-28 fusion
16	8	8.5	125	AAG45730	Arabidopsis thalia
17	8	8.5	138	ABP68966	Human polypeptide
18	8	8.5	144	AAE33860	Human retrocyclin
19	8	8.5	161	ABP68985	Human polypeptide
20	8	8.5	168	AAU31085	Novel human secret
21	8	8.5	176	AAW71696	R78-hpro-28 fusion
22	8	8.5	196	AAW71694	R21-2Hpro-28 fusio
23	8	8.5	467	AAG45700	Arabidopsis thalia
24	8	8.5	467	ABB93586	Herbicidally activ
25	8	8.5	649	AAW70259	Human concentrativ
26	8	8.5	649	AAW70260	Human concentrativ
27	8	8.5	650	AAW49107	Human concentrativ
28	8	8.5	702	AAP94014	Carcinoembryonic c
29	8	8.5	702	ABU04817	Human expressed pr
30	7	7.4	10	AAY46751	Immunogenic peptid
31	7	7.4	10	AAU96020	Carcino embryonic
32	7	7.4	14	AAB88012	CD66 peptide CD66a
33	7	7.4	14	AAB88105	CD66 peptide CD66e
34	7	7.4	15	AAY98827	HLA class II bindi
35	7	7.4	15	AAY98922	HLA class II bindi
36	7	7.4	43	AAB12435	Pro-CEMA amino aci
37	7	7.4	60	AAP91535	Modified region 30
38	7	7.4	65	AAG33208	Zea mays protein f
39	7	7.4	67	ABU51508	Helicobacter pylor
40	7	7.4	81	ABU51695	Helicobacter pylor
41	7	7.4	83	ABB97584	Novel human protei
42	7	7.4	87	AAW98778	H. pylori GHPO 118
43	7	7.4	94	AAR44818	Deduced sequence o
44	7	7.4	94	ABB79000	Human defensin 5 p
45	7	7.4	95	AAU91040	Transplant media a
46	7	7.4	96	AAU91053	Transplant media a
47	7	7.4	96	AAU91054	Transplant media a
48	7	7.4	100	AAR44819	Sequence of the ga
49	7	7.4	101	ABP76314	Human GENSET prote
50	7	7.4	105	ABB61388	Drosophila melanog
51	7	7.4	113	AAO02198	Human polypeptide
52	7	7.4	114	AAY74986	Neisseria meningit
53	7	7.4	115	AAY74984	Neisseria gonorrhe
54	7	7.4	123	AAW67855	Human secreted pro
55	7	7.4	133	AAY74985	Neisseria meningit
56	7	7.4	170	AAU51905	Propionibacterium
57	7	7.4	173	AAR22266	Collagen type prot
58	7	7.4	196	AAE10311	Tamias asiaticus h
59	7	7.4	286	AAP81224	Carcinoembryonic a
60	7	7.4	286	AAR54712	Carcinoembryonic a
61	7	7.4	286	AAW06871	Carcinoembryonic a
62	7	7.4	286	AAW83136	Human CEA protein
63	7	7.4	286	ABU04818	Human expressed pr
64	7	7.4	286	ABU04824	Human expressed pr
65	7	7.4	286	ABU04829	Human expressed pr
66	7	7.4	287	ABU04801	Human expressed pr
67	7	7.4	287	ABU04841	Human expressed pr
68	7	7.4	321	AAR06434	SP1-like protein e
69	7	7.4	321	AAR22045	Carcino embryonic
70	7	7.4	321	ABU05236	Human expressed pr
71	7	7.4	323	AAG06220	Arabidopsis thalia
72	7	7.4	323	AAG07395	Arabidopsis thalia
73	7	7.4	323	AAG48723	Arabidopsis thalia
74	7	7.4	328	AAG06219	Arabidopsis thalia
75	7	7.4	328	AAG07394	Arabidopsis thalia
76	7	7.4	328	AAG48722	Arabidopsis thalia
77	7	7.4	344	AAP81225	Carcinoembryonic a
78	7	7.4	344	AAP94000	Sequence of CEA an
79	7	7.4	344	AAR54715	Carcinoembryonic a
80	7	7.4	344	AAW06874	Carcinoembryonic a
81	7	7.4	344	AAW33139	BT20 protein. Hom
82	7	7.4	344	ABU04820	Human expressed pr

```
 83  7  7.4  344  ABU04826  Human expressed pr
 84  7  7.4  344  ABU04831  Human expressed pr
 85  7  7.4  349  ABU04803  Human expressed pr
 86  7  7.4  349  ABU04805  Human expressed pr
 87  7  7.4  349  ABU04807  Human expressed pr
 88  7  7.4  349  ABU04808  Human expressed pr
 89  7  7.4  349  ABU04835  Human expressed pr
 90  7  7.4  349  ABU04838  Human expressed pr
 91  7  7.4  351  AAR22318  Carcino embryonic
 92  7  7.4  368  AAP93396  Transmembrane (TM)
 93  7  7.4  369  AAG06218  Arabidopsis thalia
 94  7  7.4  369  AAG48721  Arabidopsis thalia
 95  7  7.4  372  ABU04802  Human expressed pr
 96  7  7.4  417  AAR22044  Carcino embryonic
 97  7  7.4  422  AAY34959  Chlamydia pneumoni
 98  7  7.4  430  AAP93994  Sequence of transm
 99  7  7.4  451  AAY37158  Protein involved i
100  7  7.4  464  AAP93995  Transmembrane (TM)
```

ALIGNMENTS

```
RESULT 1
AAY01603
ID   AAY01603 standard; protein; 94 AA.
XX
AC   AAY01603;
XX
XX
DT   18-JUN-1999  (first entry)
XX
DE   Amino acid sequence of the human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;  97FR-0010823.
XX
PR   29-AUG-1997;  97FR-0010823.
XX
PA   (GEST ) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Claim 1; Fig 1; 56pp; French.
XX
CC   The present sequence represents human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can be used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence  94 AA;

Query Match             100.0%;  Score 94;  DB 20;  Length 94;
Best Local Similarity   100.0%;  Pred. No. 2.5e-91;
Matches  94;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 MRTLTLLSAFLLVALQAWAEPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGST 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRTLTLLSAFLLVALQAWAEPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGST 60

Qy   61 KGLICHCRVLYCIFGEHLGGTCFILGERYPICCY 94
         ||||||||||||||||||||||||||||||||||
Db   61 KGLICHCRVLYCIFGEHLGGTCFILGERYPICCY 94


RESULT 2
AAY01604
ID   AAY01604 standard; protein; 94 AA.
XX
AC   AAY01604;
XX
XX
DT   18-JUN-1999  (first entry)
XX
DE   Amino acid sequence of the human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;  97FR-0010823.
XX
PR   29-AUG-1997;  97FR-0010823.
XX
PA   (GEST ) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
DR   N-PSDB; AAX26697.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Disclosure; Fig 3; 56pp; French.
XX
CC   The present sequence represents human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can be used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
```

CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 94 AA;

Query Match   100.0%;   Score 94;   DB 20;   Length 94;
Best Local Similarity   100.0%;   Pred. No. 2.5e-91;
Matches 94;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   1 MRTLTLLSAFLLVALQAWAEPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGST 60
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRTLTLLSAFLLVALQAWAEPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGST 60

Qy   61 KGLICHCRVLYCIFGEHLGGTCFILGERYPICCY 94
       ||||||||||||||||||||||||||||||||||
Db   61 KGLICHCRVLYCIFGEHLGGTCFILGERYPICCY 94

RESULT 3
AAY01606
ID   AAY01606 standard; peptide; 31 AA.
XX
AC   AAY01606;
XX
DT   18-JUN-1999 (first entry)
XX
DE   Fragment of human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;   97FR-0010823.
XX
PR   29-AUG-1997;   97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Claim 2; Page 33; 56pp; French.
XX
CC   The present sequence represents human defensin (Def-X) fragment. The
CC   Def-X polypeptide can be used as an antimicrobial, antiparasitic agent
CC   or a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating

CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence 31 AA;

Query Match   33.0%;   Score 31;   DB 20;   Length 31;
Best Local Similarity   100.0%;   Pred. No. 3.2e-25;
Matches 31;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   64 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 94
       |||||||||||||||||||||||||||||||
Db   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31

RESULT 4
ABB44491
ID   ABB44491 standard; Peptide; 58 AA.
XX
AC   ABB44491;
XX
DT   04-FEB-2002 (first entry)
XX
DE   Peptide #11997 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000;   2000US-0180312.
PR   26-MAY-2000;   2000US-0207456.
PR   30-JUN-2000;   2000US-0608408.
PR   03-AUG-2000;   2000US-0612366.
PR   21-SEP-2000;   2000US-0234687.
PR   27-SEP-2000;   2000US-0236359.
PR   04-OCT-2000;   2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 27; SEQ ID NO 37126; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a peptide encoded by a single exon
CC   nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 58 AA;

Query Match   8.5%;   Score 8;   DB 22;   Length 58;
Best Local Similarity   100.0%;   Pred. No. 1.2;
Matches 8;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   19 AEPLQARA 26
       ||||||||
Db   19 AEPLQARA 26

RESULT 5
AAU32095
ID   AAU32095 standard; Protein; 58 AA.
XX
AC   AAU32095;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Novel human secreted protein #2586.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179949-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US08656.
XX
PR   18-APR-2000; 2000US-0552929.
PR   26-JAN-2001; 2001US-0770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy -
PS   Claim 20; Page 554; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising
CC   the nucleic acids encoding the polypeptides and cells genetically
CC   engineered to express them are also useful for producing the proteins.
CC   The proteins are useful in genetic vaccination, testing and
CC   therapy, and can be used as nutritional supplements. They may be used to
CC   increase stem cell proliferation; to regulate haematopoiesis; and in
CC   bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC   immune suppression and/or stimulation; as anti-inflammatory agents; and
CC   in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC   sequences of novel human secreted proteins of the invention.
XX
SQ   Sequence 58 AA;

Query Match 8.5%; Score 8; DB 22; Length 58;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   19 AEPLQARA 26
        ||||||||
Db   19 AEPLQARA 26

RESULT 6
AAM78244
ID   AAM78244 standard; Protein; 58 AA.
XX
AC   AAM78244;
XX
DT   06-NOV-2001 (first entry)

XX
DE   Human bone marrow expressed probe encoded protein SEQ ID NO: 38550.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLS-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
PS   Example 4; SEQ ID NO: 38550; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is a
CC   protein encoded by one of the probes of the invention.
XX
SQ   Sequence 58 AA;

Query Match 8.5%; Score 8; DB 22; Length 58;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   19 AEPLQARA 26
        ||||||||
Db   19 AEPLQARA 26

RESULT 7
ABG47256
ID   ABG47256 standard; Peptide; 58 AA.
XX
AC   ABG47256;
XX
DT   19-AUG-2002 (first entry)
XX
DE   Human peptide encoded by genome-derived single exon probe SEQ ID 36921.
XX
KW   Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW   primary ciliary dyskinesis; fibrocystic pulmonary dysplasia;
KW   hyaline membrane disease.
XX
OS   Homo sapiens.
XX

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     December 17, 2003, 14:51:01 ; Search time 37.6 Seconds
                                    (without alignments)
                                    466.897 Million cell updates/sec

Title:          US-10-045-180A-3
Perfect score:  94
Sequence:       1 MRKTLTLLSAFLLIVALQAWAE.........GEHLGGTCFILGKRYPICCY 94

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       696363 seqs, 18675810 residues

Word size :     0

Total number of hits satisfying chosen parameters:       696363

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :      Published Applications_AA:*
    1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
    2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
    3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
    4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
    5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
    6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
    7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
    8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
    9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
   10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.pep:*
   11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
   12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
   13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
   14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
   15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
   16:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
   17:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
   18:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	94	100.0	94	14	US-10-013-770-3	Sequence 3, Appli
2	94	100.0	94	14	US-10-045-180A-3	Sequence 3, Appli
3	44	46.8	44	14	US-10-013-770-5	Sequence 5, Appli
4	44	46.8	44	14	US-10-045-180A-5	Sequence 6, Appli
5	31	33.0	31	14	US-10-013-770-6	Sequence 6, Appli
6	31	33.0	31	14	US-10-045-180A-6	Sequence 4, Appli
7	19	20.2	19	14	US-10-013-770-4	Sequence 7, Appli
8	19	20.2	19	14	US-10-045-180A-4	Sequence 47890, A
9	8	8.5	58	9	US-09-864-761-47890	Sequence 12, Appl
10	8	8.5	59	12	US-10-141-645-12	Sequence 12, Appl
11	8	8.5	94	14	US-10-045-180A-12	Sequence 121, App
12	8	8.5	118	12	US-10-141-645-121	Sequence 65, Appl
13	8	8.5	140	12	US-10-141-645-65	Sequence 66, Appl
14	8	8.5	140	12	US-10-141-645-66	Sequence 79, Appl
15	8	8.5	140	12	US-10-141-645-69	Sequence 69, Appl

16	8	8.5	141	12	US-10-141-645-70	Sequence 70, Appl
17	8	8.5	141	12	US-10-141-645-71	Sequence 71, Appl
18	8	8.5	141	12	US-10-141-645-123	Sequence 123, App
19	8	8.5	141	12	US-10-141-645-125	Sequence 125, App
20	7	7.4	19	12	US-10-076-802-2	Sequence 2, Appli
21	7	7.4	19	12	US-10-367-169-2	Sequence 2, Appli
22	7	7.4	87	12	US-09-882-227-422	Sequence 422, App
23	7	7.4	92	12	US-10-118-471-28	Sequence 28, Appl
24	7	7.4	92	12	US-10-029-386-31739	Sequence 31739, A
25	7	7.4	92	12	US-10-138-176-28	Sequence 28, Appl
26	7	7.4	94	12	US-10-045-180A-10	Sequence 10, Appl
27	7	7.4	95	9	US-09-917-340-76	Sequence 76, Appl
28	7	7.4	96	9	US-09-917-340-89	Sequence 89, Appl
29	7	7.4	96	9	US-09-917-340-90	Sequence 90, Appl
30	7	7.4	100	14	US-10-045-180A-11	Sequence 11, Appl
31	7	7.4	123	10	US-09-984-245-169	Sequence 169, App
32	7	7.4	123	11	US-09-966-262-169	Sequence 169, App
33	7	7.4	123	11	US-09-983-966-169	Sequence 169, App
34	7	7.4	123	15	US-10-143-090-169	Sequence 169, App
35	7	7.4	196	12	US-10-222-857-3	Sequence 3, Appli
36	7	7.4	436	15	US-10-156-761-14932	Sequence 14932, A
37	7	7.4	595	15	US-10-156-698-4871	Sequence 4871, Ap
38	7	7.4	702	15	US-10-117-937-592	Sequence 592, App
39	7	7.4	702	15	US-10-157-031-341	Sequence 341, App
40	7	7.4	702	15	US-10-207-655-87	Sequence 87, Appl
41	7	7.4	734	9	US-09-756-551A-17	Sequence 17, Appl
42	7	7.4	737	9	US-09-925-301-1133	Sequence 1133, Ap
43	7	7.4	1479	11	US-09-982-315-4	Sequence 4, Appli
44	7	7.4	1480	10	US-09-568-756-2	Sequence 2, Appli
45	7	7.4	1480	11	US-09-982-315-2	Sequence 2, Appli
46	7	7.4	1480	11	US-09-982-315-6	Sequence 6, Appli
47	7	7.4	1480	12	US-10-161-539-2	Sequence 2, Appli
48	6	6.4	10	12	US-10-026-066-72	Sequence 72, Appl
49	6	6.4	15	11	US-09-883-343A-14	Sequence 14, Appl
50	6	6.4	21	11	US-09-883-343A-13	Sequence 13, Appl
51	6	6.4	33	11	US-09-764-392-2879	Sequence 2879, Ap
52	6	6.4	61	15	US-10-156-761-13958	Sequence 13958, A
53	6	6.4	66	10	US-09-901-484A-131	Sequence 131, App
54	6	6.4	66	10	US-09-853-526-131	Sequence 131, App
55	6	6.4	68	10	US-09-901-484A-130	Sequence 130, App
56	6	6.4	68	10	US-09-853-526-130	Sequence 130, App
57	6	6.4	77	10	US-09-901-484A-125	Sequence 125, App
58	6	6.4	77	10	US-09-853-526-125	Sequence 125, App
59	6	6.4	77	15	US-10-178-213-446	Sequence 446, App
60	6	6.4	77	15	US-10-178-213-449	Sequence 449, App
61	6	6.4	80	10	US-09-893-737-12	Sequence 12, Appl
62	6	6.4	82	10	US-10-135-545A-8	Sequence 8, Appli
63	6	6.4	90	10	US-09-901-484A-129	Sequence 129, App
64	6	6.4	90	10	US-09-853-526-129	Sequence 129, App
65	6	6.4	90	11	US-09-764-872-370	Sequence 370, App
66	6	6.4	93	10	US-09-738-626-6630	Sequence 6630, Ap
67	6	6.4	97	10	US-09-901-484A-132	Sequence 132, App
68	6	6.4	97	10	US-09-853-526-132	Sequence 132, App
69	6	6.4	97	12	US-10-141-645-13	Sequence 13, Appl
70	6	6.4	97	12	US-10-311-994-23	Sequence 23, Appl
71	6	6.4	97	14	US-10-045-180A-9	Sequence 9, Appli
72	6	6.4	99	12	US-10-118-471-25	Sequence 25, Appl
73	6	6.4	99	12	US-10-138-176-25	Sequence 25, Appl
74	6	6.4	102	9	US-09-864-761-43600	Sequence 43600, A
75	6	6.4	103	9	US-09-764-869-1191	Sequence 1191, Ap
76	6	6.4	103	11	US-09-883-343A-39	Sequence 39, Appl
77	6	6.4	105	15	US-10-091-504-1191	Sequence 1191, Ap
78	6	6.4	105	12	US/10/135	Sequence 28, Appl
79	6	6.4	108	12	US/10/135	Sequence 29, Appl
80	6	6.4	111	12	US-10-135-545A-16	Sequence 16, Appl
81	6	6.4	112	12	US/10/135	Sequence 7, Appl
82	6	6.4	112	12	US-10-135-545A-10	Sequence 10, Appl
83	6	6.4	115	10	US-09-755-633-10	Sequence 86, Appl
84	6	6.4	115	10	US-10-262-439-86	Sequence 86, Appl
85	6	6.4	115	15	US-10-218-654-86	Sequence 403, App
86	6	6.4	125	11	US-09-975-719-403	Sequence 79, Appl
87	6	6.4	128	11	US-09-975-719-79	Sequence 5, Appl
88	6	6.4	134	10	US-09-755-633-5	Sequence 5, Appli

```
89   6   6.4   US-10-262-439-81     134  12   Sequence 81, Appl
90   6   6.4   US-10-218-654-81     134  15   Sequence 81, Appl
91   6   6.4   US-10-156-761-7570   137  15   Sequence 7570, Ap
92   6   6.4   US-10-080-170-351    137  16   Sequence 351, App
93   6   6.4   US-09-738-626-4179   147  10   Sequence 4179, Ap
94   6   6.4   US-09-895-298-82     154  11   Sequence 82, Appl
95   6   6.4   US-10-060-036-163    157  15   Sequence 163, App
96   6   6.4   US-09-864-761-37898  179   9   Sequence 37898, A
97   6   6.4   US-09-080-113-2      179  12   Sequence 2, Appli
98   6   6.4   US-09-901-484A-133   182  10   Sequence 133, App
99   6   6.4   US-09-853-526-133    182  10   Sequence 133, App
100  6   6.4   US-09-798-029-4      184   9   Sequence 4, Appli
```

ALIGNMENTS

RESULT 1
US-10-013-770-3
; Sequence 3, Application US/10013770
; Publication No. US20020115151A1
; GENERAL INFORMATION:
;   APPLICANT: GENSET SA
;   TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;   COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;   THERAPEUTIC APPLICATIONS
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/013,770
;   FILING DATE: 10-Dec-2001
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/486,580
;   FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
;   INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 94 AMINO ACIDS
;   TYPE: AMINO ACID
;   STRANDEDNESS: SINGLE
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE: PROTEIN
;   ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SIGNAL PEPTIDE
;   LOCATION: 1..19
;   FEATURE:
;   NAME/KEY: PRO REGION
;   LOCATION: 20..63
;   FEATURE:
;   NAME/KEY: MATURE PEPTIDE
;   LOCATION: 64..94
;   SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-013-770-3

Query Match   100.0%;   Score 94;   DB 14;   Length 94;
Best Local Similarity   100.0%;   Pred. No. 1.8e-88;
Matches   94; Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy   1  MRTLTLLSAFLLVALQAWAEPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGST 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1  MRTLTLLSAFLLVALQAWAEPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGST 60

Qy  61  KGLICHCRVLYCIFGEHLGGTCFILGERYPICCY 94
        ||||||||||||||||||||||||||||||||||
Db  61  KGLICHCRVLYCIFGEHLGGTCFILGERYPICCY 94
```

RESULT 2
US-10-045-180A-3
; Sequence 3, Application US/10045180A
; Publication No. US2002018703A1
; GENERAL INFORMATION:
;   APPLICANT: Bougueleret, Lydie
;   APPLICANT: Chumakov, Ilya
;   TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Composit[ion]
;   TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeutic
;   FILE REFERENCE: GEN-100D1
;   CURRENT APPLICATION NUMBER: US/10/045,180A
;   CURRENT FILING DATE: 2001-10-18
;   PRIOR APPLICATION NUMBER: US 09/486,580
;   PRIOR FILING DATE: 2000-02-25
;   PRIOR APPLICATION NUMBER: PCT/FR98/01864
;   PRIOR FILING DATE: 1998-08-28
;   PRIOR APPLICATION NUMBER: FR 97/10823
;   PRIOR FILING DATE: 1997-08-29
;   NUMBER OF SEQ ID NOS: 12
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 94
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(94)
;   OTHER INFORMATION: Def-X preproprotein sequence
;   FEATURE:
;   NAME/KEY: SIGNAL
;   LOCATION: (1)..(19)
;   OTHER INFORMATION: Def-X signal peptide
;   FEATURE:
;   NAME/KEY: PROPEP
;   LOCATION: (20)..(63)
;   OTHER INFORMATION: Def-X propeptide
;   FEATURE:
;   NAME/KEY: PEPTIDE
;   LOCATION: (64)..(94)
;   OTHER INFORMATION: Def-X mature peptide
US-10-045-180A-3

Query Match   100.0%;   Score 94;   DB 14;   Length 94;
Best Local Similarity   100.0%;   Pred. No. 1.8e-88;
Matches   94; Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy   1  MRTLTLLSAFLLVALQAWAEPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGST 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1  MRTLTLLSAFLLVALQAWAEPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGST 60

Qy  61  KGLICHCRVLYCIFGEHLGGTCFILGERYPICCY 94
        ||||||||||||||||||||||||||||||||||
Db  61  KGLICHCRVLYCIFGEHLGGTCFILGERYPICCY 94
```

RESULT 3
US-10-013-770-5
; Sequence 5, Application US/10013770
; Publication No. US20020115151A1
; GENERAL INFORMATION:
;   APPLICANT: GENSET SA
;   TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;   COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND

```
                    GenCore version 5.1.6
        Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   December 17, 2003, 14:50:14 ; Search time 21.056 Seconds
                                         (without alignments)
                                         188.888 Million cell updates/sec

Title:          US-10-045-180A-3
Perfect score:  94
Sequence:       1 MRTLTLLSAFLLVALQAWAE.........GEHLGGTCFILGERYPICCY 94

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       328717 seqs, 42310858 residues

Word size :     0

Total number of hits satisfying chosen parameters:   328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :   Issued Patents AA:*
   1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
   2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
   3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
   4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
   5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
   6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                          SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	94	100.0	94	4	US-09-486-580A-3	Sequence 3, Appli
2	44	46.8	44	4	US-09-486-580A-5	Sequence 5, Appli
3	31	33.0	31	4	US-09-486-580A-6	Sequence 6, Appli
4	19	20.2	19	4	US-09-486-580A-4	Sequence 4, Appli
5	8	8.5	72	3	US-09-042-071-38	Sequence 38, Appl
6	8	8.5	94	1	US-08-158-189-12	Sequence 12, Appl
7	8	8.5	94	3	US-08-158-189-13	Sequence 13, Appl
8	8	8.5	124	3	US-09-042-071-39	Sequence 39, Appl
9	8	8.5	178	3	US-09-042-071-42	Sequence 42, Appl
10	8	8.5	195	3	US-09-042-071-40	Sequence 40, Appl
11	8	8.5	649	3	US-08-800-291B-5	Sequence 5, Appli
12	8	8.5	649	3	US-08-800-291B-6	Sequence 6, Appli
13	8	8.5	650	3	US-08-800-291B-4	Sequence 4, Appli
14	7	7.4	19	4	US-09-351-657A-42	Sequence 42, Appl
15	7	7.4	19	4	US-09-351-657A-45	Sequence 45, Appl
16	7	7.4	45	4	US-09-351-657A-46	Sequence 46, Appl
17	7	7.4	58	1	US-08-158-189-2	Sequence 2, Appli
18	7	7.4	94	1	US-08-158-189-5	Sequence 5, Appli
19	7	7.4	95	1	US-08-158-189-14	Sequence 14, Appl
20	7	7.4	95	1	US-08-158-189-15	Sequence 15, Appl
21	7	7.4	100	1	US-08-158-189-7	Sequence 7, Appli
22	7	7.4	111	1	US-08-158-189-17	Sequence 17, Appl
23	7	7.4	111	1	US-08-466-886-19	Sequence 19, Appl
24	7	7.4	111	3	US-08-469-617-19	Sequence 19, Appl
25	7	7.4	284	1	US-09-252-991A-32586	Sequence 32586, A
26	7	7.4	309	1	US-08-463-092B-8	Sequence 8, Appli
27	7	7.4	309	2	US-08-460-907B-8	Sequence 8, Appli
28	7	7.4	321	6	5169835-17	Patent No. 5169835
29	7	7.4	373	4	US-09-252-991A-30414	Sequence 30414, A
30	7	7.4	422	4	US-09-198-452A-377	Sequence 377, App
31	7	7.4	464	2	US-08-602-725-32	Sequence 32, Appl
32	7	7.4	642	1	US-08-217-299-1	Sequence 1, Appli
33	7	7.4	698	2	US-08-602-725-36	Sequence 36, Appl
34	7	7.4	734	2	US-08-389-459A-17	Sequence 17, Appl
35	7	7.4	734	3	US-08-987-867A-17	Sequence 17, Appl
36	7	7.4	1476	3	US-09-256-703-2	Sequence 2, Appli
37	7	7.4	1479	4	US-08-951-912-4	Sequence 4, Appli
38	7	7.4	1479	4	US-09-174-077-4	Sequence 4, Appli
39	7	7.4	1480	1	US-07-637-521-2	Sequence 2, Appli
40	7	7.4	1480	1	US-08-136-742A-2	Sequence 2, Appli
41	7	7.4	1480	1	US-08-135-809A-3	Sequence 3, Appli
42	7	7.4	1480	1	US-08-466-886-17	Sequence 17, Appl
43	7	7.4	1480	2	US-08-951-912-2	Sequence 2, Appli
44	7	7.4	1480	2	US-08-951-912-6	Sequence 6, Appli
45	7	7.4	1480	2	US-08-469-461-2	Sequence 2, Appli
46	7	7.4	1480	2	US-08-469-461-4	Sequence 4, Appli
47	7	7.4	1480	2	US-08-691-605-2	Sequence 2, Appli
48	7	7.4	1480	2	US-08-455-552A-14	Sequence 14, Appl
49	7	7.4	1480	3	US-07-890-609-2	Sequence 2, Appli
50	7	7.4	1480	3	US-07-890-609-4	Sequence 4, Appli
51	7	7.4	1480	2	US-09-248-026-2	Sequence 2, Appli
52	7	7.4	1480	3	US-08-469-617-17	Sequence 17, Appl
53	7	7.4	1480	3	US-08-681-838A-2	Sequence 2, Appli
54	7	7.4	1480	3	US-08-681-838A-3	Sequence 3, Appli
55	7	7.4	1480	2	US-09-174-077-2	Sequence 2, Appli
56	7	7.4	1480	4	US-09-174-077-6	Sequence 6, Appli
57	7	7.4	1480	4	US-09-425-453A-2	Sequence 2, Appli
58	7	7.4	1480	4	US-09-425-453A-4	Sequence 4, Appli
59	7	7.4	1480	4	US-09-425-453A-8	Sequence 8, Appli
60	7	7.4	1480	4	US-09-425-453A-10	Sequence 10, Appl
61	7	7.4	1480	4	US-09-425-453A-12	Sequence 12, Appl
62	7	7.4	1480	4	US-09-425-453A-14	Sequence 14, Appl
63	7	7.4	1480	4	US-09-425-453A-16	Sequence 16, Appl
64	7	7.4	1480	4	US-09-425-453A-18	Sequence 18, Appl
65	7	7.4	1480	4	US-09-425-453A-20	Sequence 20, Appl
66	7	7.4	1480	5	PCT-US99-11667-2	Sequence 2, Appli
67	7	7.4	1480	5	540846-5	Patent No. 540846
68	7	7.4	1480	6	US-08-924-629C-14	Sequence 14, Appl
69	6	6.4	15	4	US-08-924-629C-13	Sequence 13, Appl
70	6	6.4	21	4	US-08-479-577-2	Sequence 2, Appli
71	6	6.4	44	2	US-08-756-977-2	Sequence 2, Appli
72	6	6.4	44	2	PCT-US93-05331-6	Sequence 6, Appli
73	6	6.4	44	5	US-09-082-358B-82	Sequence 82, Appl
74	6	6.4	58	3	US-09-338-907-131	Sequence 131, App
75	6	6.4	66	3	US-09-218-207-131	Sequence 131, App
76	6	6.4	66	3	US-09-338-907-130	Sequence 130, App
77	6	6.4	68	3	US-09-218-207-130	Sequence 130, App
78	6	6.4	68	3	US-09-338-907-125	Sequence 125, App
79	6	6.4	77	3	US-09-218-207-125	Sequence 125, App
80	6	6.4	77	4	US-09-134-001C-3127	Sequence 3127, Ap
81	6	6.4	81	3	US-09-249-542-8	Sequence 8, Appli
82	6	6.4	82	3	US-09-338-907-129	Sequence 129, App
83	6	6.4	90	3	US-09-218-207-129	Sequence 129, App
84	6	6.4	97	3	US-09-338-907-132	Sequence 132, App
85	6	6.4	97	3	US-09-309-487-23	Sequence 23, Appl
86	6	6.4	97	3	US-09-218-207-132	Sequence 132, App
87	6	6.4	103	4	US-09-967-808-23	Sequence 23, Appl
88	6	6.4	111	4	US-09-924-629C-39	Sequence 39, Appl
89	6	6.4	111	4	US-09-249-542-16	Sequence 16, Appl
90	6	6.4	112	4	US-09-107-532A-5707	Sequence 5707, Ap
91	6	6.4	115	4	US-09-249-542-7	Sequence 7, Appli
92	6	6.4	115	4	US-09-249-542-10	Sequence 10, Appl
93	6	6.4	125	4	US-09-322-409-86	Sequence 86, Appl
94	6	6.4	128	4	US-09-451-527-86	Sequence 86, Appl
95	6	6.4	134	4	US-09-199-637A-403	Sequence 403, App
96	6	6.4	134	4	US-09-199-637A-79	Sequence 79, Appl
97	6	6.4	134	4	US-09-322-409-81	Sequence 81, Appl
98	6	6.4	134	4	US-09-451-527-81	Sequence 81, Appl
99	6	6.4	134	4	US-09-371-615A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-486-580A-3
; Sequence 3, Application US/09486580A
; Patent No. 6329340
; GENERAL INFORMATION:
;   APPLICANT: GENSET SA
;   TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;   TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;   TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/486,580A
;   FILING DATE: FEBRUARY 25, 2000
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 94 AMINO ACIDS
;   TYPE: AMINO ACID
;   STRANDEDNESS: SINGLE
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE: PROTEIN
;   ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SIGNAL PEPTIDE
;   LOCATION: 1..19
;   FEATURE:
;   NAME/KEY: PRO REGION
;   LOCATION: 20..63
;   FEATURE:
;   NAME/KEY: MATURE PEPTIDE
;   LOCATION: 64..94
US-09-486-580A-3

Query Match   100.0%; Score 94; DB 4; Length 94;
Best Local Similarity 100.0%; Pred. No. 5e-92;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRTLTLLSAFLLVALQAWAEPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGST 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MRTLTLLSAFLLVALQAWAEPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGST 60

Qy 61 KGLICHCRVLYCIFGEHLGGTCFILGERYPICCY 94
       ||||||||||||||||||||||||||||||||||
Db 61 KGLICHCRVLYCIFGEHLGGTCFILGERYPICCY 94

RESULT 2
US-09-486-580A-5
; Sequence 5, Application US/09486580A
; Patent No. 6329340
; GENERAL INFORMATION:
;   APPLICANT: GENSET SA
;   TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;   TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;   TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/486,580A
;   FILING DATE: FEBRUARY 25, 2000
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 44 AMINO ACIDS
;   TYPE: AMINO ACID
;   STRANDEDNESS: SINGLE
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE: PRO REGION
;   ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
US-09-486-580A-5

Query Match   46.8%; Score 44; DB 4; Length 44;
Best Local Similarity 100.0%; Pred. No. 1.9e-39;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 EPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGSTKGL 63
       ||||||||||||||||||||||||||||||||||||||||||||
Db 1 EPLQARAHEMPAQKQPPADDQDVVIYFSGDDSCSLQVPGSTKGL 44

RESULT 3
US-09-486-580A-6
; Sequence 6, Application US/09486580A
; Patent No. 6329340
; GENERAL INFORMATION:
;   APPLICANT: GENSET SA
;   TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
;   TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
;   TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/486,580A
;   FILING DATE: FEBRUARY 25, 2000
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 6:

OM protein - protein search, using sw model

```
                 GenCore version 5.1.6
          Copyright (c) 1993 - 2003  Compugen Ltd.

Run on:    December 17, 2003, 14:50:14 ; Search time 6.696 Seconds
                                (without alignments)
                                445.225 Million cell updates/sec

Title:        US-10-045-180A-6
Perfect score: 31
Sequence:   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31

Scoring table: OLIGO
           Gapop 60.0 , Gapext 60.0

Searched:   283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters:     283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :   PIR_76:*
       1: pir1:*
       2: pir2:*
       3: pir3:*
       4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	19.4	6	115	1	G64304	hypothetical prote
2	19.4	6	124	2	H82117	hypothetical prote
3	19.4	6	131	2	AG3198	hypothetical prote
4	19.4	6	137	2	A70913	hypothetical prote
5	19.4	6	174	2	A49181	alpha-crystallin
6	19.4	6	174	2	S58758	alpha-crystallin c
7	19.4	6	212	2	AG0854	probable sugar ald
8	19.4	6	268	2	A97260	dihydropteroate sy
9	19.4	6	287	2	C90023	hypothetical prote
10	19.4	6	294	2	S19006	beta-lactamase (EC
11	19.4	6	296	2	A64110	cell division inhi
12	19.4	6	300	2	AF1286	CDP-abequose synth
13	19.4	6	300	2	AI1657	CDP-abequose synth
14	19.4	6	352	2	B87338	asparaginyl-tRNA s
15	19.4	6	356	2	B70424	lipid A disacchari
16	19.4	6	386	2	B72060	2-component regula
17	19.4	6	386	2	G86563	2-component regula
18	19.4	6	408	2	A70594	probable manA prot
19	19.4	6	410	2	F87004	probable mannose-6
20	19.4	6	456	2	A64052	glutathione-disulf
21	19.4	6	491	2	D83353	mannitol dehydroge
22	19.4	6	499	2	AH2761	penicillin-binding
23	19.4	6	506	2	G97542	penicillin-binding
24	19.4	6	510	2	A55207	glycerol-3-phospha
25	19.4	6	512	2	B83197	glycerol-3-phospha
26	19.4	6	518	2	B69981	N-acetylmuramoyl-L
27	19.4	6	543	2	T39345	probable metabolit
28	19.4	6	597	2	B53978	protein-tyrosine-p
29	19.4	6	639	2	T46577	arylsulfatase (EC
30	19.4	6	694	2	A53978	protein-tyrosine-p
31	19.4	6	716	2	C83375	probable glycosyl
32	19.4	6	813	2	T40622	translation elonga
33	19.4	6	836	2	C82726	DNA uptake protein
34	19.4	6	840	2	T02164	hypothetical prote
35	19.4	6	1061	2	C88690	protein F41H10.4 [
36	16.1	5	35	2	F70205	hypothetical prote
37	16.1	5	57	2	C95890	hypothetical prote
38	16.1	5	60	2	T00005	pepJ protein - Aci
39	16.1	5	67	2	JN0378	neurotoxin Os-3 -
40	16.1	5	80	2	PC4370	glutathione-disulf
41	16.1	5	82	2	E70670	hypothetical prote
42	16.1	5	84	2	S42603	hypothetical prote
43	16.1	5	84	2	AF0078	hypothetical prote
44	16.1	5	92	2	JH0716	neuropeptide Y pre
45	16.1	5	93	2	B83146	conserved hypothet
46	16.1	5	95	1	BORT3	prostatic steroid-
47	16.1	5	99	2	S78076	nonhistone chromos
48	16.1	5	101	2	D99233	hypothetical prote
49	16.1	5	101	2	S59321	hypothetical prote
50	16.1	5	102	2	T49503	hypothetical prote
51	16.1	5	102	2	AE2503	hypothetical prote
52	16.1	5	110	2	S09950	hypothetical prote
53	16.1	5	111	2	T36555	probable membrane
54	16.1	5	119	2	G65203	yijD protein - Esc
55	16.1	5	119	2	G91240	hypothetical prote
56	16.1	5	119	2	D86088	hypothetical prote
57	16.1	5	119	2	AD0935	probable membrane
58	16.1	5	119	2	S64079	probable membrane
59	16.1	5	120	2	S21565	hypothetical prote
60	16.1	5	121	2	H71053	hypothetical prote
61	16.1	5	125	1	CCQFCP	cytochrome c' - Rh
62	16.1	5	126	4	CCQFCR	cytochrome c' - Rh
63	16.1	5	128	1	CCRFCX	cytochrome c' - Rh
64	16.1	5	129	1	CCRFPP	cytochrome c' - Rh
65	16.1	5	129	2	H87262	conserved hypothet
66	16.1	5	130	2	S34493	cellulase (EC 3.2.
67	16.1	5	131	1	T10769	profilin - Para ru
68	16.1	5	133	2	B69279	conserved hypothet
69	16.1	5	134	1	HSMS34	histone H3.4 - mou
70	16.1	5	135	2	H71941	hypothetical prote
71	16.1	5	135	2	A64653	hypothetical prote
72	16.1	5	137	2	B83475	hypothetical prote
73	16.1	5	142	2	D75495	two component resp
74	16.1	5	145	2	AF2820	hypothetical prote
75	16.1	5	145	2	F97598	actinorhodin polyk
76	16.1	5	149	2	C72410	hypothetical cytos
77	16.1	5	149	2	AF3343	conserved hypothet
78	16.1	5	152	2	AE2708	conserved hypothet
79	16.1	5	152	2	F97490	conserved hypothet
80	16.1	5	157	2	AF2708	hypothetical prote
81	16.1	5	157	2	G97490	conserved hypothet
82	16.1	5	160	2	A71300	conserved hypothet
83	16.1	5	161	2	B89812	disease resistance
84	16.1	5	164	2	T07759	calcium channel be
85	16.1	5	164	2	S62212	hypothetical prote
86	16.1	5	165	2	S73194	alpha-crystallin B
87	16.1	5	172	2	I48171	hypothetical prote
88	16.1	5	172	2	T26837	ribosomal protein
89	16.1	5	173	1	S13068	NADH dehydrogenase
90	16.1	5	173	2	B90613	NADH dehydrogenase
91	16.1	5	173	2	B90615	NADH dehydrogenase
92	16.1	5	173	2	B90621	NADH dehydrogenase
93	16.1	5	173	2	B90625	hypothetical prote
94	16.1	5	174	2	B83390	conserved hypothet
95	16.1	5	175	2	F90205	alpha-crystallin c
96	16.1	5	175	1	CYBOAB	alpha-crystallin c
97	16.1	5	175	1	CYHUAB	alpha-crystallin c
98	16.1	5	175	2	A53871	alpha-crystallin c
99	16.1	5	175	2	A39608	alpha-crystallin c
100	16.1	5	175	2	I53319	alpha B-crystallin

ALIGNMENTS

RESULT 1
G64304
hypothetical protein MJ0039 - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2003
C;Accession: G64304
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: G64304
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-115 <BUL>
A;Cross-references: GB:U67462; GB:L77117; NID:g1590835; PIDN:AAB98020.1; PID:g1590840; T
C;Genetics:
A;Map position: REV39998-39651
A;Start codon: TTG
C;Superfamily: RNA polymerase, subunit F

Query Match   19.4%;   Score 6;   DB 1;   Length 115;
Best Local Similarity   100.0%;   Pred. No. 17;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   21 ILGERY 26
          ||||||
Db   14 ILGERY 19

RESULT 2
H82117
hypothetical protein VC2105 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: H82117
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: H82117
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-124 <HEI>
A;Cross-references: GB:AE004284; GB:AE003852; NID:g9655649; PIDN:AAF95251.1; GSPDB:GN001
C;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC2105
A;Map position: 1

Query Match   19.4%;   Score 6;   DB 2;   Length 124;
Best Local Similarity   100.0%;   Pred. No. 18;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   21 ILGERY 26
          ||||||
Db   78 ILGERY 83

RESULT 3
AG3198
hypothetical protein Atu5317 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 18-Nov-2002
C;Accession: AG3198
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, l

erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AG3198
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-131 <KUR>
A;Cross-references: GB:AE008687; PIDN:AAL46005.1; PID:g17743760; GSPDB:GN00188
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu5317
A;Genome: plasmid

Query Match   19.4%;   Score 6;   DB 2;   Length 131;
Best Local Similarity   100.0%;   Pred. No. 19;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   12 GEHLGG 17
          ||||||
Db   52 GEHLGG 57

RESULT 4
A70913
hypothetical protein Rv0049 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70913
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.B.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70913
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-137 <COL>
A;Cross-references: GB:Z80775; GB:AL123456; NID:g3250715; PIDN:CAB02528.1; PID:g1568588
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0049
C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0049

Query Match   19.4%;   Score 6;   DB 2;   Length 137;
Best Local Similarity   100.0%;   Pred. No. 20;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   11 FGEHLG 16
          ||||||
Db   61 FGEHLG 66

RESULT 5
A49181
alpha B-crystallin - chicken
C;Species: Gallus gallus (chicken)
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C;Accession: A49181
R;Sawada, K.; Agata, K.; Eguchi, G.
Exp. Eye Res. 55, 879-887, 1992
A;Title: Crystallin gene expression in the process of lentoidogenesis in cultures of ch
A;Reference number: A49181; MUID:93137981; PMID:1283129
A;Accession: A49181
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-174 <SAW>
A;Cross-references: GB:S53164; NID:g263999; PIDN:AAB25041.1; PID:g264000

```
                    GenCore version 5.1.6
         Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   December 17, 2003, 14:50:14 ; Search time 5.208 seconds
                                (without alignments)
                        279.921 Million cell updates/sec

Title:          US-10-045-180A-6
Perfect score:  1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31
Sequence:       1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       127863 seqs, 47026705 residues

Word size :     0

Total number of hits satisfying chosen parameters:      127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :      SwissProt_41:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                            SUMMARIES
```

Result No.	Score	* Query Match	Length	DB	ID	Description
1	19.4	6	52	1	CRAB_COLLI	O12987 columba liv
2	19.4	6	52	1	CRAB_TURME	O12995 turdus meru
3	19.4	6	115	1	Y039_METJA	Q60351 methanococc
4	19.4	6	137	1	Y049_MYCTU	P71706 mycobacteri
5	19.4	6	174	1	CRAB_ANAPL	Q05557 anas platyr
6	19.4	6	174	1	CRAB_CHICK	Q05713 gallus gall
7	19.4	6	294	1	BLAC_CITDI	P22390 citrobacter
8	19.4	6	296	1	YFCH_HAEIN	P71373 haemophilus
9	19.4	6	356	1	LPXB_AQUAE	O67420 aquifex aeo
10	19.4	6	456	1	GSHR_HAEIN	P43783 haemophilus
11	19.4	6	512	1	GLPD_PSEAE	P52111 pseudomonas
12	19.4	6	813	1	EFG2_SCHPO	O94429 schizosacch
13	19.4	6	4655	1	LRP2_HUMAN	P98164 homo sapien
14	16.1	5	68	1	PHYB_ASPFI	P81440 aspergillus
15	16.1	5	35	1	Y847_BORBU	O51787 borrelia bu
16	16.1	5	52	1	CRAB_BUDEL	O12991 eudromia el
17	16.1	5	52	1	CRAB_TRASC	Q91518 trachemys s
18	16.1	5	67	1	SCX3_ORTSC	P15225 orthochirus
19	16.1	5	84	1	YO74_RHIME	P42878 rhizobium m
20	16.1	5	93	1	VE7_HPV50	Q80928 human papil
21	16.1	5	95	1	PSC3_RAT	P02780 rattus norv
22	16.1	5	96	1	CU85_HUMAN	P59088 homo sapien
23	16.1	5	99	1	NHPB_YEAST	P11633 saccharomyc
24	16.1	5	110	1	US36_HCMVA	P16842 human cytom
25	16.1	5	119	1	YGH2_YEAST	P53161 saccharomyc
26	16.1	5	119	1	YIJD_ECOLI	P27308 escherichia
27	16.1	5	121	1	YIPP_DROME	Q9xzf0 drosophila
28	16.1	5	125	1	CYCP_RHOPH	P00145 rhodospiril
29	16.1	5	126	1	CYCP_RHORU	P00144 rhodospiril
30	16.1	5	128	1	CYCP_RHOSP	P00146 rhodopseudo
31	16.1	5	129	1	CYCP_RHOCA	P00147 rhodobacter
32	16.1	5	130	1	GUN2_PERAE	P23666 persea amer
33	16.1	5	131	1	PRO1_HEVBR	O65812 hevea brasi

34	16.1	5	131	1	PRO1_MALDO	Q9xf40 malus domes
35	16.1	5	131	1	PRO3_HEVBR	Q9m7n0 hevea brasi
36	16.1	5	131	1	PRO5_HEVBR	Q9m7m8 hevea brasi
37	16.1	5	131	1	PRO6_HEVBR	Q91ei8 hevea brasi
38	16.1	5	131	1	PROF_PYRCO	Q9xf38 pyrus commu
39	16.1	5	134	1	H34_MOUSE	P02301 mus musculu
40	16.1	5	134	1	PRO1_OLEEU	O24169 olea europa
41	16.1	5	134	1	PRO2_OLEEU	O24170 olea europa
42	16.1	5	134	1	PRO3_OLEEU	O24171 olea europa
43	16.1	5	150.	1	PADI_RHILO	Q9eyul rhizobium 1
44	16.1	5	150	1	PAD2_RHILO	Q982g0 rhizobium 1
45	16.1	5	160	1	Y650_TREPA	O83656 treponema p
46	16.1	5	165	1	YC36_PORPU	P51273 porphyra pu
47	16.1	5	170	1	CT29_MOUSE	Q9d742 mus musculu
48	16.1	5	171	1	ALL8_OLEEU	Q9m7r0 olea europa
49	16.1	5	173	1	NU6M_STRCA	O21407 struthio ca
50	16.1	5	173	1	RL10_SYNY3	P23350 synechocyst
51	16.1	5	175	1	CRAB_BOVIN	P02510 bos taurus
52	16.1	5	175	1	CRAB_HUMAN	P02511 homo sapien
53	16.1	5	175	1	CRAB_MESAU	P05811 mesocricetu
54	16.1	5	175	1	CRAB_MOUSE	P23927 mus musculu
55	16.1	5	175	1	CRAB_RABIT	P41316 oryctolagus
56	16.1	5	175	1	CRAB_RAT	P23928 rattus norv
57	16.1	5	175	1	Y773_METJA	Q58183 methanococc
58	16.1	5	181	1	RM06_ACACA	P46765 acanthamoeb
59	16.1	5	183	1	TPA_BPT4	P03739 bacteriopha
60	16.1	5	192	1	COMK_BACSU	P40396 bacillus su
61	16.1	5	192	1	MOBA_PYRHO	O58708 pyrococcus
62	16.1	5	193	1	AX2D_PHAAU	O24542 phaseolus a
63	16.1	5	194	1	CSR3_HUMAN	P50461 homo sapien
64	16.1	5	194	1	CSR3_MOUSE	P50462 mus musculu
65	16.1	5	194	1	CSR3_RAT	P50463 rattus norv
66	16.1	5	198	1	DENR_HUMAN	O43583 homo sapien
67	16.1	5	198	1	DENR_MOUSE	Q9cqj6 mus musculu
68	16.1	5	200	1	CT29_HUMAN	Q9nus5 homo sapien
69	16.1	5	206	1	RL3_BORBU	P94267 borrelia bu
70	16.1	5	210	1	YGEK_ECO57	P58664 escherichia
71	16.1	5	210	1	YGEK_ECOLI	Q46791 escherichia
72	16.1	5	218	1	GT27_SCHMA	P35661 schistosoma
73	16.1	5	220	1	FGF3_CHICK	P48801 gallus gall
74	16.1	5	224	1	SNG2_MOUSE	O55101 mus musculu
75	16.1	5	226	1	RB3C_BOVIN	P10949 bos taurus
76	16.1	5	229	1	LIPB_XYLFA	Q9pdv9 xylella fas
77	16.1	5	234	1	SNG2_RAT	O54980 rattus norv
78	16.1	5	235	1	TIRA_HUMAN	P58753 homo sapien
79	16.1	5	236	1	28KD_MYCLE	P19361 mycobacteri
80	16.1	5	237	1	60IM_PROMI	P22833 proteus mir
81	16.1	5	237	1	FGF3_XENLA	P36386 xenopus lae
82	16.1	5	239	1	FGF3_HUMAN	P11487 homo sapien
83	16.1	5	241	1	ISPD_MYCLE	Q9ccw6 mycobacteri
84	16.1	5	245	1	FGF3_MOUSE	P05524 mus musculu
85	16.1	5	246	1	DAPB_CHLMU	Q9pk30 chlamydia m
86	16.1	5	256	1	FGF3_BRARE	P48802 brachydanio
87	16.1	5	257	1	EUTC_RHOER	Q59782 rhodococcus
88	16.1	5	261	1	TFPB_MORBO	P20666 moraxella b
89	16.1	5	263	1	AMPM_SALTY	P10882 salmonella
90	16.1	5	267	1	BLAC_SERFO	P80545 serratia fo
91	16.1	5	269	1	DAPF_CHLMU	Q9pjw2 chlamydia m
92	16.1	5	273	1	NPSN_DROME	Q9vxk0 drosophila
93	16.1	5	275	1	DAPF_CHLTR	O84437 chlamydia t
94	16.1	5	286	1	HEMK_VIBCH	Q9kq26 vibrio chol
95	16.1	5	291	1	KPPP_ALCEU	P19924 alcaligenes
96	16.1	5	291	1	KPPR_ALCEU	P19923 alcaligenes
97	16.1	5	294	1	YGHA_ECOLI	P25887 escherichia
98	16.1	5	295	1	RSP4_BOVIN	P26452 bos taurus
99	16.1	5	295	1	RSP4_CRIGR	P38982 cricetulus
100	16.1	5	295	1	RSP4_HUMAN	P08865 homo sapien

```
                                              ALIGNMENTS


RESULT 1
```

```
CRAB_COLLI
ID   CRAB_COLLI      STANDARD;       PRT;    52 AA.
AC   O12987;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Alpha crystallin B chain (Alpha(B)-crystallin) (Fragment).
GN   CRYAB.
OS   Columba livia (Domestic pigeon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Columbiformes; Columbidae; Columba.
OX   NCBI_TaxID=8932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97271643; PubMed=9126559;
RA   Caspers G.J., Uit de Weerd D., Wattel J., de Jong W.W.;
RT   "Alpha-crystallin sequences support a galliform/anseriform clade.";
RL   Mol. Phylogenet. Evol. 7:185-188(1997).
CC   -!- FUNCTION: MAY CONTRIBUTE TO THE TRANSPARENCY AND REFRACTIVE INDEX
CC       OF THE LENS.
CC   -!- SUBUNIT: AGGREGATES WITH HOMOLOGOUS PROTEINS, INCLUDING ALPHA-A-
CC       CRYSTALLIN AND THE SMALL HEAT SHOCK PROTEIN HSP28, TO FORM LARGE
CC       HETEROMERIC COMPLEXES.
CC   -!- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC       FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X96594; CAA65412.1; -.
DR   InterPro; IPR003090; Crystallin_N.
DR   InterPro; IPR002068; Hsp20.
DR   Pfam; PF00525; crystallin; 1.
DR   ProDom; PD001193; Crystallin_N; 1.
DR   PROSITE; PS01031; HSP20; PARTIAL.
KW   Eye lens protein.
FT   NON_TER       1      1
FT   NON_TER      52     52
SQ   SEQUENCE   52 AA;   6078 MW;  B4FC7ECF2B79FE5 CRC64;

Query Match             19.4%;   Score 6;  DB 1;  Length 52;
Best Local Similarity  100.0%;   Pred. No. 5.7;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 IFGEHL 15
           ||||||
Db      19 IFGEHL 24


RESULT 2
CRAB_TURME
ID   CRAB_TURME      STANDARD;       PRT;    52 AA.
AC   O12995;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Alpha crystallin B chain (Alpha(B)-crystallin) (Fragment).
GN   CRYAB.
OS   Turdus merula (Blackbird).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Turdidae; Turdus.
OX   NCBI_TaxID=9187;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97271643; PubMed=9126559;
RA   Caspers G.J., Uit de Weerd D., Wattel J., de Jong W.W.;
RT   "Alpha-crystallin sequences support a galliform/anseriform clade.";
RL   Mol. Phylogenet. Evol. 7:185-188(1997).
CC   -!- FUNCTION: MAY CONTRIBUTE TO THE TRANSPARENCY AND REFRACTIVE INDEX
CC       OF THE LENS.
CC   -!- SUBUNIT: AGGREGATES WITH HOMOLOGOUS PROTEINS, INCLUDING ALPHA-A-
CC       CRYSTALLIN AND THE SMALL HEAT SHOCK PROTEIN HSP28, TO FORM LARGE
CC       HETEROMERIC COMPLEXES.
CC   -!- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC       FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is not removed. Usage
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X96595; CAA65413.1; -.
DR   InterPro; IPR003090; Crystallin_N.
DR   InterPro; IPR002068; Hsp20.
DR   Pfam; PF00525; crystallin; 1.
DR   ProDom; PD001193; Crystallin_N; 1.
DR   PROSITE; PS01031; HSP20; PARTIAL.
KW   Eye lens protein.
FT   NON_TER       1      1
FT   NON_TER      52     52
SQ   SEQUENCE   52 AA;   6143 MW;  82271ECE2619EF6F CRC64;

Query Match             19.4%;   Score 6;  DB 1;  Length 52;
Best Local Similarity  100.0%;   Pred. No. 5.7;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 IFGEHL 15
           ||||||
Db      19 IFGEHL 24


RESULT 3
Y039_METJA
ID   Y039_METJA      STANDARD;       PRT;   115 AA.
AC   Q60351;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ0039.
GN   MJ0039.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
```

```
                  GenCore version 5.1.6
          Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    December 17, 2003, 14:50:14 ; Search time 14.632 Seconds
                                          (without alignments)
                              546.722 Million cell updates/sec

Title:          US-10-045-180A-6
Perfect score:  31
Sequence:       1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31
Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       830525 seqs, 258052604 residues

Word size :     0

Total number of hits satisfying chosen parameters:    830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :       SPTREMBL_23:*
          1:  sp_archea:*
          2:  sp_bacteria:*
          3:  sp_fungi:*
          4:  sp_human:*
          5:  sp_invertebrate:*
          6:  sp_mammal:*
          7:  sp_mhc:*
          8:  sp_organelle:*
          9:  sp_phage:*
         10:  sp_plant:*
         11:  sp_rodent:*
         12:  sp_virus:*
         13:  sp_vertebrate:*
         14:  sp_unclassified:*
         15:  sp_rvirus:*
         16:  sp_bacteriap:*
         17:  sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	19.4	6	39	13	Q8JJ30	Q8jj30 ficedula al
2	19.4	6	39	13	Q8JJ31	Q8jj31 ficedula hy
3	19.4	6	124	16	Q9KQ95	Q9kq95 vibrio chol
4	19.4	6	131	16	Q8UK07	Q8uk07 agrobacteri
5	19.4	6	136	2	O69489	O69489 mycobacteri
6	19.4	6	191	8	Q8WDD5	Q8wdd5 taenia soli
7	19.4	6	212	16	Q8XER3	Q8xer3 salmonella
8	19.4	6	233	5	Q8MUQ6	Q8muq6 anopheles g
9	19.4	6	236	13	Q90Z77	Q90z77 brachydanio
10	19.4	6	250	5	Q9VA75	Q9va75 drosophila
11	19.4	6	268	16	Q97F29	Q97f29 clostridium
12	19.4	6	287	16	Q99S15	Q99s15 staphylococ
13	19.4	6	294	2	Q46029	Q46029 citrobacter
14	19.4	6	296	8	Q95813	Q95813 hymenolepis
15	19.4	6	296	8	Q8M675	Q8m675 paragonimus
16	19.4	6	296	8	Q8HN32	Q8hn32 paragonimus
17	19.4	6	297	8	Q9B8W5	Q9b8w5 taenia cras
18	19.4	6	297	8	Q8M6K9	Q8m6k9 taenia soli
19	19.4	6	300	8	Q9B8Y4	Q9b8y4 fasciola he
20	19.4	6	300	8	Q34521	Q34521 fasciola he
21	19.4	6	300	16	Q92AW6	Q92aw6 listeria in
22	19.4	6	300	16	Q8Y6J0	Q8y6j0 listeria mo
23	19.4	6	309	11	Q8VGP0	Q8vgp0 mus musculu
24	19.4	6	313	3	O59869	O59869 orpinomyces
25	19.4	6	323	17	Q8TWC8	Q8twc8 methanopyru
26	19.4	6	350	16	Q98IZ0	Q98iz0 rhizobium l
27	19.4	6	352	16	Q9AA86	Q9aa86 caulobacter
28	19.4	6	355	16	Q9RJU4	Q9rju4 streptomyce
29	19.4	6	359	16	Q989G1	Q989g1 rhizobium l
30	19.4	6	386	16	Q927W8	Q927w8 chlamydia p
31	19.4	6	394	16	Q9K2D1	Q9k2d1 chlamydia p
32	19.4	6	396	16	Q8EQ89	Q8eq89 oceanobacil
33	19.4	6	408	16	Q8KRQ4	Q8krq4 mycobacteri
34	19.4	6	408	16	O05898	O05898 mycobacteri
35	19.4	6	410	6	Q8WMT0	Q8wmt0 megadezma l
36	19.4	6	410	16	Q9CCJ5	Q9ccj5 mycobacteri
37	19.4	6	412	11	Q8BSQ6	Q8bsq6 mus musculu
38	19.4	6	419	16	Q8CJT2	Q8cjt2 streptomyce
39	19.4	6	423	16	Q8KLW9	Q8klw9 bathyergus
40	19.4	6	423	17	Q8ZU85	Q8zu85 pyrobaculum
41	19.4	6	424	11	Q8K1V2	Q8k1v2 echimys chr
42	19.4	6	424	11	Q8K1N9	Q8k1n9 thryonomys
43	19.4	6	424	11	Q8K1Q9	Q8k1q9 petromus ty
44	19.4	6	452	5	Q25631	Q25631 onchocerca
45	19.4	6	452	11	Q9CUL7	Q9cul7 mus musculu
46	19.4	6	459	5	Q9BI89	Q9bi89 caenorhabdi
47	19.4	6	462	5	O01412	O01412 onchocerca
48	19.4	6	468	16	Q98C98	Q98c98 rhizobium l
49	19.4	6	491	16	Q91ID6	Q91id6 pseudomonas
50	19.4	6	506	16	Q8UF94	Q8uf94 agrobacteri
51	19.4	6	513	13	Q918V9	Q918v9 raja eglant
52	19.4	6	514	3	Q9FB48	Q9fb48 corynebacte
53	19.4	6	518	16	O32041	O32041 bacillus su
54	19.4	6	523	16	Q8EVP4	Q8evp4 mycoplasma
55	19.4	6	528	5	Q9VGT3	Q9vgt3 drosophila
56	19.4	6	543	3	O94342	O94342 schizosacch
57	19.4	6	551	17	Q8TVW6	Q8tvw6 methanopyru
58	19.4	6	597	13	Q91871	Q91871 xenopus lae
59	19.4	6	599	16	Q8PFA1	Q8pfa1 xanthomonas
60	19.4	6	599	16	Q8P3S5	Q8p3s5 xanthomonas
61	19.4	6	639	3	O43113	O43113 neurospora
62	19.4	6	694	13	Q91870	Q91870 xenopus lae
63	19.4	6	716	16	Q91IV5	Q91iv5 pseudomonas
64	19.4	6	836	16	Q9PEF0	Q9pef0 xylella fas
65	19.4	6	840	10	O64804	O64804 arabidopsis
66	19.4	6	862	6	Q9BE52	Q9be52 macaca fasc
67	19.4	6	873	16	Q9S024	Q9s024 escherichia
68	19.4	6	1018	1	Q8NKK1	Q8nkk1 acidianus a
69	19.4	6	1170	4	Q8NE11	Q8ne11 homo sapien
70	19.4	6	1170	11	Q8CDL4	Q8cdl4 mus musculu
71	19.4	6	1170	11	Q8BND3	Q8bnd3 mus musculu
72	19.4	6	1219	4	Q9P2L0	Q9p2l0 homo sapien
73	19.4	6	2635	12	O40942	O40942 kaposi's sa
74	19.4	6	2635	12	P88955	P88955 kaposi's sa
75	16.1	5	46	9	Q38295	Q38295 lactococcus
76	16.1	5	50	16	Q9QTZ3	Q9qtz3 tt virus. o
77	16.1	5	52	16	Q8F683	Q8f683 leptospira
78	16.1	5	54	10	Q8GYC8	Q8gyc8 arabidopsis
79	16.1	5	57	16	Q92WF2	Q92wf2 rhizobium m
80	16.1	5	57	17	Q96X71	Q96x71 sulfolobus
81	16.1	5	59	2	Q9KJD7	Q9kjd7 hafnia alve
82	16.1	5	60	1	O57699	O57699 acidianus a
83	16.1	5	61	3	P89500	P89500 saccharomyc
84	16.1	5	62	3	Q96V76	Q96v76 caionectria
85	16.1	5	62	16	Q8FE45	Q8fe45 escherichia
86	16.1	5	63	9	Q8HAM2	Q8ham2 burkholderi
87	16.1	5	64	11	Q91Y98	Q91y98 mus musculu
88	16.1	5	66	16	Q93519	Q93519 salmonella
89	16.1	5	68	2	Q05625	Q05625 staphylococ

```
 90   5   16.1   70    3   Q8X1M4    Q8x1m4 calonectria
 91   5   16.1   71    3   Q8X1M2    Q8x1m2 calonectria
 92   5   16.1   74   10   Q93VW1    Q93vw1 oryza sativ
 93   5   16.1   77    7   O43207    O43207 homo sapien
 94   5   16.1   77    9   Q8H9Y3    Q8H9y3 pseudomonas
 95   5   16.1   77   10   Q9S955    Q9s955 mesembryant
 96   5   16.1   77   12   Q9DHU7    Q9dhu7 yaba-like d
 97   5   16.1   82    8   Q94UZ4    Q94uz4 gracula rel
 98   5   16.1   83    8   Q8SKE8    Q8ske8 gracula rel
 99   5   16.1   84   16   Q8ZI77    Q8zi77 yersinia pe
100   5   16.1   85    3   Q8X1M0    Q8x1m0 calonectria
```

ALIGNMENTS

RESULT 1
Q8JJ30
ID   Q8JJ30   PRELIMINARY;   PRT;   39 AA.
AC   Q8JJ30;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Alpha-B-crystallin (Fragment).
GN   ABC.
OS   Ficedula albicollis.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Muscicapidae; Ficedula.
OX   NCBI_TaxID=59894;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B8;
RX   MEDLINE=21918460; PubMed=11918793;
RA   Primmer C.R., Borge T., Lindell J., Saetre G.-P.;
RT   "Single-nucleotide polymorphism characterization in species with
RT   limited available sequence information: high nucleotide diversity
RT   revealed in the avian genome.";
RL   Mol. Ecol. 11:603-612(2002).
DR   EMBL; AF454243; AAM22922.1; -.
DR   InterPro; IPR003090; Crystallin_N.
DR   Pfam; PF00525; crystallin; 1.
DR   ProDom; PD001193; Crystallin_N; 1.
FT   NON_TER   1   1
FT   NON_TER   39   39
SQ   SEQUENCE   39 AA;   4442 MW;   24034FE4C5BD1E55 CRC64;

Query Match   19.4%; Score 6; DB 13; Length 39;
Best Local Similarity   100.0%; Pred. No. 15;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   10 IFGEHL 15
               ||||||
Db   12 IFGEHL 17

RESULT 2
Q8JJ31
ID   Q8JJ31   PRELIMINARY;   PRT;   39 AA.
AC   Q8JJ31;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Alpha-B-crystallin (Fragment).
GN   ABC.
OS   Ficedula hypoleuca.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Muscicapidae; Ficedula.
OX   NCBI_TaxID=46689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sp1;
RX   MEDLINE=21918460; PubMed=11918793;
RA   Primmer C.R., Borge T., Lindell J., Saetre G.-P.;
RT   "Single-nucleotide polymorphism characterization in species with
RT   limited available sequence information: high nucleotide diversity
RT   revealed in the avian genome.";
RL   Mol. Ecol. 11:603-612(2002).
DR   EMBL; AF454242; AAM22921.1; -.
DR   InterPro; IPR003090; Crystallin_N.
DR   Pfam; PF00525; crystallin; 1.
DR   ProDom; PD001193; Crystallin_N; 1.
FT   NON_TER   1   1
FT   NON_TER   39   39
SQ   SEQUENCE   39 AA;   4442 MW;   24034FE4C5BD1E55 CRC64;

Query Match   19.4%; Score 6; DB 13; Length 39;
Best Local Similarity   100.0%; Pred. No. 15;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   10 IFGEHL 15
               ||||||
Db   12 IFGEHL 17

RESULT 3
Q9KQ95
ID   Q9KQ95   PRELIMINARY;   PRT;   124 AA.
AC   Q9KQ95;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical protein VC2105.
GN   VC2105.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004284; AAF95251.1; -.
DR   TIGR; VC2105; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   124 AA;   14590 MW;   6E68B67B976CBBF3 CRC64;

Query Match   19.4%; Score 6; DB 16; Pred. No. 44;
Best Local Similarity   100.0%;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   21 ILGERY 26
               ||||||
Db   78 ILGERY 83

RESULT 4
Q8UK07
ID   Q8UK07   PRELIMINARY;   PRT;   131 AA.
AC   Q8UK07;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Atu5317.
GN   ATU5317.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OG   Plasmid AT.

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: December 17, 2003, 14:50:14 ; Search time 18.352 Seconds
(without alignments)
268.119 Million cell updates/sec

Title: US-10-045-180A-6
Perfect score: 31
Sequence: 1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31
Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 1107863 seqs, 158726573 residues
Word size : 0
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	31	100.0	31	20	AAY01606	Fragment of human
2	31	100.0	94	20	AAY01603	Amino acid sequenc
3	31	100.0	94	20	AAY01604	Amino acid sequenc
4	6	19.4	37	23	ABP26644	Streptococcus poly
5	6	19.4	64	22	AAU58954	Propionibacterium
6	6	19.4	81	21	AAG59901	Arabidopsis thalia
7	6	19.4	81	21	AAG61227	Arabidopsis thalia
8	6	19.4	111	22	AAU57814	Propionibacterium
9	6	19.4	137	23	ABU05700	M. tuberculosis an
10	6	19.4	179	22	ABG52130	Human liver peptid
11	6	19.4	179	22	ABB32059	Peptide #4710 enco
12	6	19.4	179	22	ABB37307	Peptide #4813 enco
13	6	19.4	179	22	ABB22600	Protein #4599 enco
14	6	19.4	179	22	AAM58003	Human brain expres
15	6	19.4	179	22	AAM70444	Human bone marrow
16	6	19.4	179	22	AAM18276	Peptide #4710 enco
17	6	19.4	179	22	AAM30765	Peptide #4802 enco
18	6	19.4	179	22	AAM05888	Peptide #4570 enco
19	6	19.4	179	23	ABG40078	Human peptide enco
20	6	19.4	249	23	AAU91178	Pantothenate kinas
21	6	19.4	275	22	ABB70744	Drosophila melanog
22	6	19.4	294	22	ABG06705	Novel human diagno
23	6	19.4	296	22	AAU35611	Haemophilus influe
24	6	19.4	300	23	ABB49371	Listeria monocytog
25	6	19.4	337	20	AAY35212	Chlamydia pneumoni
26	6	19.4	368	22	ABB04596	Lettuce big-vein v
27	6	19.4	408	22	AAG81217	Mycobacterium tube
28	6	19.4	456	23	AAB88530	Haemophilus influe
29	6	19.4	456	22	AAU91461	Haemophilus influe
30	6	19.4	498	22	ABG24816	Novel human diagno
31	6	19.4	499	22	ABG24685	Novel human diagno
32	6	19.4	512	24	ABJ18765	Pseudomonas aerugi
33	6	19.4	528	22	ABB67113	Drosophila melanog
34	6	19.4	545	23	AAU75783	Human protein phos
35	6	19.4	716	22	ABJ18808	Pseudomonas aerugi
36	6	19.4	4655	17	AAR97208	Human calcium sens
37	6	19.4	4655	17	AAR97209	Human placental ca
38	6	19.4	4655	17	AAR97210	Human kidney calci
39	6	19.4	4655	17	AAR97211	Human parathyroid
40	6	19.4	4655	19	AAW43311	Human calcium sens
41	6	19.4	4655	19	AAW43312	Human placental ca
42	6	19.4	4655	19	AAW43313	Human kidney calci
43	6	19.4	4655	19	AAW43314	Human parathyroid
44	6	19.4	4655	24	ABP56836	Human megalin prot
45	6	19.4	4689	23	ABG04530	Novel human diagno
46	5	16.1	13	15	AAR46829	pH 2.5 acid phosph
47	5	16.1	19	19	ABP46151	Human BLyS binding
48	5	16.1	20	23	AAB65927	Human secreted pro
49	5	16.1	21	22	AAU89614	Insulin/insulin-li
50	5	16.1	22	22	AAY01451	Secreted protein e
51	5	16.1	24	24	AAY12811	Human 5' EST secre
52	5	16.1	24	24	ABP35824	LDHC_XENLA positio
53	5	16.1	27	20	AAY27707	Human secreted pro
54	5	16.1	40	21	AAB5168	Human pancreatic c
55	5	16.1	46	22	ABP35780	Ddcdc2 positions p
56	5	16.1	52	22	ABB03954	Human musculoskele
57	5	16.1	52	23	ABP34118	Human ORF3091 prot
58	5	16.1	52	24	ABP34254	Human glycoprotein
59	5	16.1	53	22	ABU13248	Novel human muscul
60	5	16.1	53	23	ABP31305	Human ORF278 prote
61	5	16.1	55	23	ABP32567	Human ORF1540 prot
62	5	16.1	55	22	ABP01122	Human ORFX protein
63	5	16.1	56	22	ABG52884	Human liver peptid
64	5	16.1	56	22	AAU59558	Propionibacterium
65	5	16.1	56	22	AAU66153	Propionibacterium
66	5	16.1	56	22	ABG02704	Novel human diagno
67	5	16.1	56	22	ABB38036	Peptide #5542 enco
68	5	16.1	56	22	AAM58666	Human brain expres
69	5	16.1	56	22	AAM71172	Human bone marrow
70	5	16.1	56	22	AAM31453	Peptide #5490 enco
71	5	16.1	56	23	ABG40969	Human peptide enco
72	5	16.1	59	22	AAM80038	Human immune/haema
73	5	16.1	65	22	ABB16086	Human nervous syst
74	5	16.1	66	21	AAB57014	Human prostate can
75	5	16.1	67	22	AAU45709	Propionibacterium
76	5	16.1	67	23	ABG63264	Human prostate spe
77	5	16.1	68	23	ABP03828	Human ORFX protein
78	5	16.1	69	23	ABP42696	Human ovarian anti
79	5	16.1	70	22	AAM84241	Human immune/haema
80	5	16.1	71	22	ABB69664	Drosophila melanog
81	5	16.1	71	22	AAU49208	Propionibacterium
82	5	16.1	71	23	ABP05221	Human ORFX protein

```
83    5   16.1   AAU52524   72   22   Propionibacterium
84    5   16.1   AAU57734   72   22   Propionibacterium
85    5   16.1   AAU40466   73   22   Propionibacterium
86    5   16.1   AAM82541   73   22   Human immune/haema
87    5   16.1   ABP04083   73   23   Human ORFX protein
88    5   16.1   AAU50025   75   22   Propionibacterium
89    5   16.1   ABB16158   75   22   Human nervous syst
90    5   16.1   ABP35207   77   23   Human transcriptio
91    5   16.1   AAU46902   78   23   Propionibacterium
92    5   16.1   AAM80786   78   22   Human haematologic
93    5   16.1   AAM81231   78   22   Human haematologic
94    5   16.1   AAM81596   78   22   Human haematologic
95    5   16.1   AAM81841   78   22   Human haematologic
96    5   16.1   ABU04778   78   24   Human expressed pr
97    5   16.1   ABU04779   78   24   Human expressed pr
98    5   16.1   ABU04780   78   24   Human expressed pr
99    5   16.1   ABU04781   78   24   Human expressed pr
100   5   16.1   AAU46468   79   22   Propionibacterium
```

ALIGNMENTS

RESULT 1
AAY01606
ID   AAY01606 standard; peptide; 31 AA.
XX
AC   AAY01606;
XX
DT   18-JUN-1999 (first entry)
XX
DE   Fragment of human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;   97FR-0010823.
XX
PR   29-AUG-1997;   97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Claim 2; Page 33; 56pp; French.
XX
CC   The present sequence represents human defensin (Def-X) fragment. The
CC   Def-X polypeptide can be used as an antimicrobial, antiparasitic agent
CC   or a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence   31 AA;

Query Match   100.0%; Score 31; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.1e-24;
Matches   31; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31
              |||||||||||||||||||||||||||||||
Db   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31

RESULT 2
AAY01603
ID   AAY01603 standard; protein; 94 AA.
XX
AC   AAY01603;
XX
DT   18-JUN-1999 (first entry)
XX
DE   Amino acid sequence of the human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;   97FR-0010823.
XX
PR   29-AUG-1997;   97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Claim 1; Fig 1; 56pp; French.
XX
CC   The present sequence represents human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX

SQ   Sequence   94 AA;

Query Match   100.0%;   Score 31;   DB 20;   Length 94;
Best Local Similarity 100.0%;   Pred. No. 3e-24;
Matches   31;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31
            ||||||||||||||||||||||||||||||||
Db   64 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 94

RESULT 3
AAY01604
ID   AAY01604 standard; protein; 94 AA.
XX
AC   AAY01604;
XX
DT   18-JUN-1999 (first entry)
XX
DE   Amino acid sequence of the human defensin (Def-X) protein.
XX
KW   Human defensin; Def-X; antimicrobial; antiparasitic; pesticide;
KW   cytostatic; anticancer; inflammation; tissue repair;
KW   endocrine regulation; corticostatic regulation; cancer; melanoma;
KW   AIDS; immune deficiency; psoriasis.
XX
OS   Homo sapiens.
XX
PN   FR2767832-A1.
XX
PD   05-MAR-1999.
XX
PF   29-AUG-1997;   97FR-0010823.
XX
PR   29-AUG-1997;   97FR-0010823.
XX
PA   (GEST) GENSET SA.
XX
PI   Bougueleret L, Chumakov I;
XX
DR   WPI; 1999-183266/16.
DR   N-PSDB; AAX26697.
XX
PT   Human defensin polypeptide Def-X - useful as antimicrobial agent,
PT   anticancer agent, pesticide, etc.
XX
PS   Disclosure; Fig 3; 56pp; French.
XX
CC   The present sequence represents human defensin (Def-X). The Def-X
CC   polypeptide can be used as an antimicrobial, antiparasitic agent or
CC   a pesticide. The Def-X polypeptide can be used as a cytostatic
CC   (especially anticancer) agent, and as an agent for modulating processes
CC   of inflammation, tissue repair and endocrine (especially corticostatic)
CC   regulation. The polypeptide can be used in a composition for external
CC   topical use, especially in a cosmetic composition. Compositions
CC   containing Def-X can be used for prevention and treatment of microbial
CC   and parasitic infections, especially where the microbial or parasitic
CC   infections are Gram-positive or -negative bacterial infections or
CC   mycobacterial, fungal or spirochaete infections, or where the viral
CC   infections are associated with enveloped viruses, especially HSV and HIV.
CC   The compositions can used for prevention and/or treatment of cancers,
CC   especially melanomas, or liver cancer, prostate cancer, non-small-cell
CC   lung cancer or colorectal carcinoma, and for enhancing immunity,
CC   especially in the case of AIDS, or preventing immune deficiency,
CC   especially in the treatment of psoriasis, as well as for modulating
CC   inflammatory processes, especially in the case of chronic inflammatory
CC   disorders.
XX
SQ   Sequence   94 AA;

Query Match   100.0%;   Score 31;   DB 20;   Length 94;
Best Local Similarity 100.0%;   Pred. No. 3e-24;
Matches   31;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31
            ||||||||||||||||||||||||||||||||
Db   64 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 94

RESULT 4
ABP26644
ID   ABP26644 standard; Protein; 37 AA.
XX
AC   ABP26644;
XX
DT   02-JUL-2002 (first entry)
XX
DE   Streptococcus polypeptide SEQ ID NO 2464.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001;   2001WO-GB04789.
XX
PR   27-OCT-2000;   2000GB-0026333.
PR   24-NOV-2000;   2000GB-0028727.
PR   07-MAR-2001;   2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   N-PSDB; ABN67275.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein -
XX
PS   Claim 1; Page 3396; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence   37 AA;

Query Match   19.4%;   Score 6;   DB 23;   Length 37;
Best Local Similarity 100.0%;   Pred. No. 39;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   21 ILGERY 26
            ||||||

Db   7 ILGERY 12

RESULT 5
AAU58954
ID   AAU58954 standard; Protein; 64 AA.
XX
AC   AAU58954;
XX
DT   13-FEB-2002 (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #19850.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW, Persing DH, Mitcham JL, Wang SS,   Bhatia A;
PI   L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59596.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 20149; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particulary involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 64 AA;

Query Match   19.4%; Score 6; DB 22; Length 64;
Best Local Similarity   100.0%; Pred. No. 64;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   14 HLGGTC 19
                 ||||||

Db   44 HLGGTC 49

RESULT 6
AAG59901
ID   AAG59901 standard; Protein; 81 AA.
XX
AC   AAG59901;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 77532.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   21-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   25-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139119.
PR   16-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.
PR   17-JUN-1999; 99US-0139492.
PR   18-JUN-1999; 99US-0139454.
PR   18-JUN-1999; 99US-0139455.

```
GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 17, 2003, 14:51:01 ; Search time 12.4 Seconds
                                             (without alignments)
                                  466.897 Million cell updates/sec

Title:          US-10-045-180A-6
Perfect score:  31
Sequence:       1 ICHCRVLYCIFGEHLGGTCPILGERYPICCY 31

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       696363 seqs, 186758610 residues

Word size :     0

Total number of hits satisfying chosen parameters:        696363

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :      Published Applications AA.:*
    1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
    2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
    3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
    4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
    5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
    6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
    7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
    8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
    9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
   10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
   11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
   12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
   13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
   14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
   15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
   16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
   17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
   18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	100.0	31	14	US-10-013-770-6	Sequence 6, Appli
2	31	100.0	31	14	US-10-045-180A-6	Sequence 6, Appli
3	31	100.0	94	14	US-10-013-770-3	Sequence 3, Appli
4	31	100.0	94	14	US-10-045-180A-3	Sequence 351, Appl
5	6	19.4	137	16	US-10-080-170-351	Sequence 37898, A
6	6	19.4	179	9	US-09-864-761-37898	Sequence 61, Appl
7	6	19.4	249	10	US-09-813-453A-61	Sequence 11204, A
8	6	19.4	296	9	US-09-815-242-11204	Sequence 268, App
9	6	19.4	408	10	US-09-712-363-268	Sequence 78, Appl
10	6	19.4	456	15	US-10-260-877-78	Sequence 114, App
11	6	19.4	512	15	US-10-127-032-114	Sequence 157, App
12	6	19.4	716	15	US-10-127-032-157	Sequence 371, App
13	5	16.1	7	11	US-09-954-385-371	Sequence 2162, Ap
14	5	16.1	19	11	US-09-880-748-2162	Sequence 151, App
15	5	16.1	21	11	US-09-776-724A-151	
16	5	16.1	27	11	US-09-974-879-311	Sequence 311, App
17	5	16.1	27	11	US-09-305-736-311	Sequence 311, App
18	5	16.1	27	12	US-09-818-683-311	Sequence 311, App
19	5	16.1	40	9	US-09-925-297-620	Sequence 620, App
20	5	16.1	50	15	US-10-178-213-462	Sequence 462, App
21	5	16.1	52	10	US-09-764-877-1901	Sequence 1901, Ap
22	5	16.1	56	9	US-09-864-761-43359	Sequence 43359, A
23	5	16.1	66	10	US-09-925-300-1592	Sequence 1592, Ap
24	5	16.1	67	12	US-10-012-952A-193	Sequence 193, App
25	5	16.1	70	10	US-09-843-676-190	Sequence 190, App
26	5	16.1	70	10	US-09-438-486-190	Sequence 190, App
27	5	16.1	70	15	US-10-053-758-190	Sequence 190, App
28	5	16.1	70	15	US-10-054-295-190	Sequence 190, App
29	5	16.1	70	15	US-10-054-611-190	Sequence 190, App
30	5	16.1	75	12	US-10-029-386-29202	Sequence 29202, A
31	5	16.1	78	10	US-09-796-692-1150	Sequence 1150, Ap
32	5	16.1	78	10	US-09-796-692-1595	Sequence 1595, Ap
33	5	16.1	78	10	US-09-796-692-1960	Sequence 1960, Ap
34	5	16.1	78	10	US-09-796-692-2205	Sequence 2205, Ap
35	5	16.1	78	15	US-10-040-862-1150	Sequence 1150, Ap
36	5	16.1	78	15	US-10-040-862-1595	Sequence 1595, Ap
37	5	16.1	78	15	US-10-040-862-1960	Sequence 1960, Ap
38	5	16.1	78	15	US-10-040-862-2205	Sequence 2205, App
39	5	16.1	79	9	US-09-764-887-235	Sequence 235, App
40	5	16.1	79	15	US-10-073-961-235	Sequence 235, App
41	5	16.1	87	15	US-09-796-692-2287	Sequence 2287, Ap
42	5	16.1	87	15	US-10-040-862-2287	Sequence 2287, Ap
43	5	16.1	92	11	US-09-764-891-4802	Sequence 4802, Ap
44	5	16.1	92	15	US-10-091-572-357	Sequence 357, App
45	5	16.1	94	15	US-09-864-761-37684	Sequence 37684, App
46	5	16.1	95	10	US-09-934-054-12	Sequence 12, Appl
47	5	16.1	95	15	US-09-985-911-27	Sequence 27, Appl
48	5	16.1	100	15	US-10-178-213-461	Sequence 461, App
49	5	16.1	101	10	US-09-738-626-5294	Sequence 5294, Ap
50	5	16.1	113	15	US-09-764-891-2931	Sequence 2931, Ap
51	5	16.1	123	9	US-09-815-242-5611	Sequence 5611, Ap
52	5	16.1	123	9	US-09-815-242-12665	Sequence 12665, A
53	5	16.1	125	11	US-09-815-242-28706	Sequence 28706, A
54	5	16.1	125	15	US-09-764-891-4868	Sequence 4868, Ap
55	5	16.1	128	11	US-10-205-428-393	Sequence 393, App
56	5	16.1	131	15	US-09-815-242-4873	Sequence 4873, Ap
57	5	16.1	137	15	US-10-106-698-6931	Sequence 6931, Ap
58	5	16.1	138	15	US-10-156-761-14880	Sequence 14880, A
59	5	16.1	144	9	US-09-815-242-9116	Sequence 9116, Ap
60	5	16.1	148	15	US-09-815-242-10517	Sequence 10517, A
61	5	16.1	148	11	US-09-764-853-674	Sequence 674, App
62	5	16.1	148	15	US-09-764-898-294	Sequence 294, App
63	5	16.1	148	15	US-09-764-881-91	Sequence 91, App
64	5	16.1	159	14	US-10-073-865-113	Sequence 113, App
65	5	16.1	167	12	US-10-227-629-10	Sequence 10, Appl
66	5	16.1	176	12	US-10-062-254-342	Sequence 342, App
67	5	16.1	177	14	US-10-032-631C-50	Sequence 50, Appl
68	5	16.1	183	13	US-10-032-201B-308	Sequence 308, App
69	5	16.1	190	13	US-10-105-427-2	Sequence 2, Appli
70	5	16.1	190	9	US-10-016-447-15	Sequence 15, Appli
71	5	16.1	194	12	US-10-016-447-4	Sequence 4, Appl
72	5	16.1	194	14	US-09-773-926-17	Sequence 17, Appl
73	5	16.1	194	15	US-10-291-172-224	Sequence 224, App
74	5	16.1	196	10	US-10-126-099-8	Sequence 8, Appli
75	5	16.1	198	12	US-10-219-220-265	Sequence 265, App
76	5	16.1	198	15	US-09-738-626-5436	Sequence 5436, Ap
77	5	16.1	200	12	US-10-346-000A-4	Sequence 4, Appli
78	5	16.1	202	12	US-09-291-172-600	Sequence 600, App
79	5	16.1	205	9	US-10-219-220-275	Sequence 275, App
80	5	16.1	209	15	US-09-738-769A-2	Sequence 2, Appli
81	5	16.1	210	15	US-10-017-161-1332	Sequence 1332, Ap
82	5	16.1	216	12	US-10-156-761-12216	Sequence 12216, A
83	5	16.1	216	12	US-10-032-585-7190	Sequence 7190, Ap
84	5	16.1	218	14	US-10-141-531-7	Sequence 7, Appli
85	5	16.1	222	12	US-10-024-599-6	Sequence 6, Appli
86	5	16.1	224	12	US-10-094-749-2215	Sequence 2215, Ap
87	5	16.1	228	10	US-09-755-665-37	Sequence 37, Appl
88	5	16.1	228	10	US-09-738-626-4341	Sequence 151, App

```
89    5    16.1    235    10    US-09-860-696-2        Sequence 2, Appli
90    5    16.1    235    15    US-10-188-947-3        Sequence 3, Appli
91    5    16.1    236    16    US-10-080-170-21       Sequence 21, Appl
92    5    16.1    238    15    US-10-156-761-13747    Sequence 13747, A
93    5    16.1    239     9    US-09-822-485-6        Sequence 6, Appli
94    5    16.1    239     9    US-09-750-963-8        Sequence 8, Appli
95    5    16.1    239     9    US-09-902-773A-9       Sequence 9, Appli
96    5    16.1    239     9    US-09-425-021-11       Sequence 11, Appl
97    5    16.1    239    10    US-09-995-598-62       Sequence 62, Appl
98    5    16.1    239    11    US-09-345-373-21       Sequence 21, Appl
99    5    16.1    239    12    US-10-189-360-12       Sequence 12, Appl
100   5    16.1    239    12    US-10-374-207-6        Sequence 6, Appli
```

ALIGNMENTS

RESULT 1
US-10-013-770-6
; Sequence 6, Application US/10013770
; Publication No. US20020115151A1
; GENERAL INFORMATION:
; APPLICANT: GENSET SA
; TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAc
; COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
; THERAPEUTIC APPLICATIONS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 550 West C Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/013,770
; FILING DATE: 10-Dec-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/486,580
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hart, Daniel
; REGISTRATION NUMBER: 40,637
; REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: MATURE PEPTIDE
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-013-770-6

Query Match   100.0%;   Score 31; DB 14; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.3e-25;
Matches   31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31
          |||||||||||||||||||||||||||||||
Db   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31

RESULT 2
US-10-045-180A-6
; Sequence 6, Application US/10045180A
; Publication No. US20020182703A1
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Human Defensin Polypeptide Def-X, Genomic DNA and cDNA, Composit
; TITLE OF INVENTION: Containing Them and Applications to Diagnosis and to Therapeuti
; FILE REFERENCE: GEN-100D1
; CURRENT APPLICATION NUMBER: US/10/045,180A
; CURRENT FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 09/486,580
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: PCT/FR98/01864
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: FR 97/10823
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(31)
; OTHER INFORMATION: Def-X mature peptide
US-10-045-180A-6

Query Match   100.0%;   Score 31; DB 14; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.3e-25;
Matches   31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31
          |||||||||||||||||||||||||||||||
Db   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31

RESULT 3
US-10-013-770-3
; Sequence 3, Application US/10013770
; Publication No. US20020115151A1
; GENERAL INFORMATION:
; APPLICANT: GENSET SA
; TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
; COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
; THERAPEUTIC APPLICATIONS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 550 West C Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/013,770
; FILING DATE: 10-Dec-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/486,580
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hart, Daniel
; REGISTRATION NUMBER: 40,637
; REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 94 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
```

GenCore version 5.1.6

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 14:50:14 ; Search time 6.944 Seconds
(without alignments)
188.888 Million cell updates/sec

Title: US-10-045-180A-6
Perfect score: 31
Sequence: 1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues
Word size :   0
Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :   Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	100.0	31	4	US-09-486-580A-6	Sequence 6, Appli
2	31	100.0	94	4	US-09-486-580A-3	Sequence 3, Appli
3	6	19.4	111	4	US-09-107-532A-5707	Sequence 5707, Ap
4	6	19.4	136	4	US-09-252-991A-22070	Sequence 22070, A
5	6	19.4	147	4	US-09-252-991A-31557	Sequence 31557, A
6	6	19.4	323	4	US-09-252-991A-24673	Sequence 24673, A
7	6	19.4	337	4	US-09-198-452A-630	Sequence 630, App
8	6	19.4	417	4	US-09-252-991A-28991	Sequence 28991, A
9	6	19.4	429	4	US-09-252-991A-20363	Sequence 20363, A
10	6	19.4	443	4	US-09-252-991A-31911	Sequence 31911, A
11	6	19.4	530	4	US-09-252-991A-18090	Sequence 18090, A
12	6	19.4	535	4	US-09-252-991A-28410	Sequence 28410, A
13	6	19.4	581	4	US-09-252-991A-25414	Sequence 25414, A
14	6	19.4	652	4	US-09-252-991A-30238	Sequence 30238, A
15	6	19.4	4654	3	US-08-476-515A-84	Sequence 84, Appl
16	6	19.4	4655	3	US-08-652-877-84	Sequence 84, Appl
17	6	19.4	4655	3	US-08-652-877-86	Sequence 86, Appl
18	6	19.4	4655	3	US-08-652-877-88	Sequence 88, Appl
19	6	19.4	4655	3	US-08-652-877-90	Sequence 90, Appl
20	5	16.1	13	1	US-07-923-724-57	Sequence 57, Appl
21	5	16.1	13	2	US-08-609-426A-57	Sequence 57, Appl
22	5	16.1	13	2	US-08-374-652C-44	Sequence 44, Appl
23	5	16.1	18	2	US-08-374-652C-84	Sequence 84, Appl
24	5	16.1	18	2	US-08-374-652C-87	Sequence 87, Appl
25	5	16.1	18	2	US-08-374-652C-88	Sequence 88, Appl
26	5	16.1	70	3	US-08-851-843A-190	Sequence 190, App
27	5	16.1	70	3	US-08-974-549A-309	Sequence 309, App
28	5	16.1	70	3	US-08-854-050-190	Sequence 190, App
29	5	16.1	70	3	US-09-430-323-190	Sequence 190, App
30	5	16.1	75	6	5175383-2	Patent No. 5175383
31	5	16.1	95	1	US-08-455-896-7	Sequence 7, Appli
32	5	16.1	95	2	US-09-933-149-7	Sequence 7, Appli
33	5	16.1	95	3	US-09-082-343-7	Sequence 7, Appli
34	5	16.1	95	3	US-09-082-253-7	Sequence 7, Appli
35	5	16.1	95	3	US-08-821-451A-27	Sequence 27, Appl
36	5	16.1	95	4	US-09-263-810-27	Sequence 27, Appl
37	5	16.1	95	4	US-09-583-169-27	Sequence 27, Appl
38	5	16.1	95	4	US-09-162-622-7	Sequence 7, Appli
39	5	16.1	95	5	PCT-US96-08235-7	Sequence 7, Appli
40	5	16.1	112	5	US-08-311-731A-22	Sequence 22, Appl
41	5	16.1	115	4	US-09-252-991A-20440	Sequence 20440, A
42	5	16.1	149	4	US-09-252-991A-21985	Sequence 21985, A
43	5	16.1	157	1	US-09-252-991A-22237	Sequence 22237, A
44	5	16.1	159	1	US-08-597-236-9	Sequence 9, Appli
45	5	16.1	159	1	US-08-746-682A-9	Sequence 9, Appli
46	5	16.1	173	4	US-09-732-210-538	Sequence 538, App
47	5	16.1	183	1	US-08-441-629-15	Sequence 15, Appl
48	5	16.1	183	3	US-08-776-207-15	Sequence 15, Appl
49	5	16.1	183	3	US-08-507-773-15	Sequence 15, Appl
50	5	16.1	183	5	PCT-US95-09172-15	Sequence 15, Appl
51	5	16.1	190	1	US-08-441-629-4	Sequence 4, Appli
52	5	16.1	190	3	US-08-776-207-4	Sequence 4, Appli
53	5	16.1	190	3	US-09-507-773-4	Sequence 4, Appli
54	5	16.1	190	5	PCT-US95-09172-4	Sequence 4, Appli
55	5	16.1	193	1	US-09-252-991A-22938	Sequence 22938, A
56	5	16.1	194	2	US-08-616-368A-11	Sequence 11, Appl
57	5	16.1	194	3	US-08-739-485-8	Sequence 8, Appli
58	5	16.1	194	4	US-09-054-298-11	Sequence 11, Appl
59	5	16.1	194	4	US-08-818-655-11	Sequence 11, Appl
60	5	16.1	194	4	US-09-305-839-11	Sequence 11, Appl
61	5	16.1	197	4	US-08-198-452A-271	Sequence 271, App
62	5	16.1	205	4	US-09-107-532A-5214	Sequence 5214, Ap
63	5	16.1	225	4	US-09-107-532A-7302	Sequence 7302, Ap
64	5	16.1	228	4	US-09-252-991A-18371	Sequence 18371, A
65	5	16.1	228	4	US-09-252-991A-28331	Sequence 28331, A
66	5	16.1	231	4	US-09-252-991A-31713	Sequence 31713, A
67	5	16.1	232	4	US-09-733-622C-31	Sequence 31, Appl
68	5	16.1	233	3	US-09-328-352-7602	Sequence 7602, Ap
69	5	16.1	236	3	US-08-711-417C-174	Sequence 174, App
70	5	16.1	236	3	US-08-711-417C-175	Sequence 175, App
71	5	16.1	236	3	US-08-793-701-56	Sequence 56, Appl
72	5	16.1	236	3	US-08-793-701-58	Sequence 58, Appl
73	5	16.1	236	4	US-09-579-264-56	Sequence 56, Appl
74	5	16.1	236	4	US-09-579-264-58	Sequence 58, Appl
75	5	16.1	237	4	US-09-107-532A-6583	Sequence 6583, Ap
76	5	16.1	238	3	US-08-711-417C-176	Sequence 176, App
77	5	16.1	238	4	US-08-711-417C-177	Sequence 177, App
78	5	16.1	239	3	US-08-464-590A-11	Sequence 11, Appl
79	5	16.1	239	1	US-08-462-169B-11	Sequence 11, Appl
80	5	16.1	239	4	US-08-207-412B-14	Sequence 14, Appl
81	5	16.1	239	2	US-08-951-822-35	Sequence 35, Appl
82	5	16.1	239	3	US-09-103-079-11	Sequence 11, Appl
83	5	16.1	239	3	US-09-023-082A-21	Sequence 21, Appl
84	5	16.1	239	4	US-09-093-585-11	Sequence 11, Appl
85	5	16.1	239	3	US-09-368-951-35	Sequence 35, Appl
86	5	16.1	239	4	US-09-425-021-11	Sequence 11, Appl
87	5	16.1	239	4	US-08-449-249-12	Sequence 12, Appl
88	5	16.1	239	4	US-08-390-207-18	Sequence 18, Appl
89	5	16.1	239	4	US-09-229-947-35	Sequence 35, Appl
90	5	16.1	239	4	US-09-417-721-7	Sequence 7, Appli
91	5	16.1	240	6	5175383-7	Patent No. 5175383
92	5	16.1	241	2	US-08-439-725A-11	Sequence 11, Appl
93	5	16.1	245	2	US-08-867-471-11	Sequence 11, Appl
94	5	16.1	245	2	US-08-438-439C-7	Sequence 7, Appli
95	5	16.1	245	3	US-08-438-439C-17	Sequence 17, Appl
96	5	16.1	245	3	US-08-705-245-12	Sequence 12, Appl
97	5	16.1	245	4	US-08-374-652C-88	Sequence 88, Appl
98	5	16.1	245	4	US-09-390-207-28	Sequence 28, Appl
99	5	16.1	246	4	US-09-252-991A-20602	Sequence 20602, A
100	5	16.1	246	4	US-09-252-991A-22967	Sequence 22967, A

ALIGNMENTS

RESULT 1
US-09-486-580A-6
; Sequence 6, Application US/09486580A
; Patent No. 6329340
; GENERAL INFORMATION:
; APPLICANT: GENSET SA
; TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
; TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
; TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/486,580A
;   FILING DATE: FEBRUARY 25, 2000
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 31 AMINO ACIDS
;   TYPE: AMINO ACID
;   STRANDEDNESS: SINGLE
;   TOPOLOGY: LINEAR
; MOLECULE TYPE: MATURE PEPTIDE
; ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
US-09-486-580A-6

Query Match   100.0%; Score 31; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.2e-24;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31
        |||||||||||||||||||||||||||||||
Db   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31

RESULT 2
US-09-486-580A-3
; Sequence 3, Application US/09486580A
; Patent No. 6329340
; GENERAL INFORMATION:
; APPLICANT: GENSET SA
; TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC
; TITLE OF INVENTION: COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND
; TITLE OF INVENTION: THERAPEUTIC APPLICATIONS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe, Martens, Olson & Bear
;   STREET: 550 West C Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/486,580A
;   FILING DATE: FEBRUARY 25, 2000
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hart, Daniel
;   REGISTRATION NUMBER: 40,637
;   REFERENCE/DOCKET NUMBER: GENSET.064C1
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 94 AMINO ACIDS
;   TYPE: AMINO ACID
;   STRANDEDNESS: SINGLE
;   TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
; FEATURE:
;   NAME/KEY: SIGNAL PEPTIDE
;   LOCATION: 1..19
; FEATURE:
;   NAME/KEY: PRO REGION
;   LOCATION: 20..63
; FEATURE:
;   NAME/KEY: MATURE PEPTIDE
;   LOCATION: 64..94
US-09-486-580A-3

Query Match   100.0%; Score 31; DB 4; Length 94;
Best Local Similarity 100.0%; Pred. No. 3e-24;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 31
        |||||||||||||||||||||||||||||||
Db   64 ICHCRVLYCIFGEHLGGTCFILGERYPICCY 94

RESULT 3
US-09-107-532A-5707
; Sequence 5707, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;   ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;   STREET: 100 Beaver Street
;   CITY: Waltham
;   STATE: Massachusetts
;   COUNTRY: USA
;   ZIP: 02354
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: CD/ROM ISO9660
;   COMPUTER: PC
;   OPERATING SYSTEM: <Unknown>
;   SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/107,532A
;   FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/085,598
;   FILING DATE: 14 May 1998
;   APPLICATION NUMBER: 60/051571
;   FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Ariniello, Pamela Deneke
;   REGISTRATION NUMBER: 40,489
;   REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION: